N° 81-552-MIF au catalogue — No.18
ISSN: 1480-9524
ISBN: 978-0-662-07659-9

Enquête internationale sur l'alphabétisation des adultes

Littératie et marché du travail :

formation de compétences et incidence sur les gains de la population de souche

par David A. Green et W. Craig Riddell

Division de la Culture, tourisme et centre de la statistique de l'éducation
Immeuble principal, Pièce 2001, Ottawa, K1A 0T6

Téléphone : 1-800-307-3382 Télécopieur : 1-613-951-9040

Statistique Canada Statistics Canada

Canada

Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.ca. Vous pouvez également communiquer avec nous par courriel à infostat@statcan.ca ou par téléphone entre 8h et 16h30 du lundi au vendredi aux :

Numéros sans frais (Canada et États-Unis) :	
Service de renseignements	1-800-263-1136
Service national d'appareils de télécommunications pour les malentendants	1-800-363-7629
Télécopieur	1-877-287-4369
Renseignements concernant le Programme des services de dépôt	1-800-635-7943
Télécopieur pour le Programme des services de dépôt	1-800-565-7757
Centre de renseignements de Statistique Canada :	1-613-951-8116
Télécopieur	1-613-951-0581

Renseignements pour accéder au produit

Le produit nº 89-552-MIF au catalogue est disponible gratuitement sous format électronique. Pour obtenir un exemplaire, il suffit de visiter notre site Web à www.statcan.ca et de choisir la rubrique Publications.

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle qui sont observées par les employés lorsqu'ils offrent des services à la clientèle. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur www.statcan.ca sous À propos de nous > Offrir des services aux Canadiens.

2008-00020/

Enquête internationale sur l'alphabétisation des adultes

Littératie et marché du travail :

formation de compétences et incidence sur les gains de la population de souche

par David A. Green et W. Craig Riddell
Département d'économique, Université de Colombie-Britannique

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'industrie, 2007

Tous droits réservés. Le contenu de la présente publication électronique peut être reproduit en tout ou en partie, et par quelque moyen que ce soit, sans autre permission de Statistique Canada, sous réserve que la reproduction soit effectuée uniquement à des fins d'étude privée, de recherche, de critique, de compte rendu ou en vue d'en préparer un résumé destiné aux journaux et/ou à des fins non commerciales. Statistique Canada doit être cité comme suit : Source (ou « Adapté de », s'il y a lieu) : Statistique Canada, année de publication, nom du produit, numéro au catalogue, volume et numéro, période de référence et page(s). Autrement, il est interdit de reproduire le contenu de la présente publication, ou de l'emmagasiner dans un système d'extraction, ou de le transmettre sous quelque forme ou par quelque moyen que ce soit, reproduction électronique, mécanique, photographique, pour quelque fin que ce soit, sans l'autorisation écrite préalable des Services d'octroi de licences, Division des services à la clientèle, Statistique Canada, Ottawa, Ontario, Canada K1A 0T6.

Novembre 2007

N° 89-552-MIF au catalogue, n° 18
ISSN 1480-9524
ISBN 978-0-662-07659-9
Périodicité : irrégulier
Ottawa

This publication is available in English (Catalogue no. 89-552-XIE, no. 18).

Statistique Canada

Acronymes

EIAA	l'Enquête internationale sur l'alphabétisation des adultes
EIACA	l'Enquête international sur l'alphabétisation et les competences des adultes
ELNEJ	Enquête longitudinale nationale sur les enfants et les jeunes
EPA	l'Enquête sur la population active
MCO	Moindres carrés ordinaires

Remerciements

Le rapport a été rendu possible grâce au soutien financier de Statistique Canada. Nous tenons à remercier Scott Murray, qui nous a encouragés à entreprendre ce projet, ainsi que les nombreux réviseurs qui nous ont fait part de leurs commentaires : Yvan Clermont et Scott Murray de Statistique Canada; Urvashi Dhawan-Biswal et Chris Li de l'équipe de la Recherche nationale en politiques sur l'apprentissage de Ressources humaines et Développement social Canada; et les réviseurs de Citoyenneté et Immigration Canada. Nous remercions également Danielle Baum de Statistique Canada de sa collaboration au rapport.

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises, les administrations canadiennes et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.

Table des matières

Acronymes		4
Remerciements		5
1.	Introduction	8
2.	Données	10
3.	Répartition de la littératie	12
4.	Formation des compétences en littératie	15
	4.1 Incidence de la scolarisation et des antécédents	15
	4.2 Effets d'âge et de génération	20
	4.3 Formation de compétences en littératie dans le milieu de travail après les études	33
5.	Incidence de la littératie sur les gains	36
	5.1 Un simple cadre théorique	36
	5.2 Résultats d'estimation	39
6.	Conclusion	43
Référence		44
Annexe A	Définitions de la scolarisation dans l'EIAA 1994, l'EIACA 2003 et l'EPA	45
Annexe B	Résultats du premier degré pour les années de scolarité	47
Notes en fin de texte		48

Table des matières

Liste des graphiques

Graphique 1	Les résultats à l'échelle des textes schématiques, 2003	12
Graphique 2	La compréhension de textes schématiques, âgée de 26 à 35 ans	21
Graphique 3	La compréhension de textes schématiques, âgée de 36 à 45 ans	22
Graphique 4	La compréhension de textes schématiques, âgée de 46 à 55 ans	22
Graphique 5	La compréhension de textes schématiques, âgée de 56 à 65 ans	23
Graphique 6	La compréhension de textes schématiques, moins que des études secondaires	24
Graphique 7	La compréhension de textes schématiques, études secondaires	24
Graphique 8	La compréhension de textes schématiques, études postsecondaires non universitaires	25
Graphique 9	La compréhension de textes schématiques, université	25
Graphique 10	La compréhension de textes schématiques, âgée de 26 à 35 en 1994	26
Graphique 11	La compréhension de textes schématiques, âgée de 36 à 45 en 1994	26
Graphique 12	La compréhension de textes schématiques, âgée de 46 à 55 en 1994	27
Graphique 13	La compréhension de textes schématiques, âgée de 56 à 65 en 1994	27
Graphique 14	La compréhension de textes schématiques, moins que des études secondaires, âgée de 26 à 45 en 1994	28
Graphique 15	La compréhension de textes schématiques, études secondaires, âgée de 26 à 45 en 1994	29
Graphique 16	La compréhension de textes schématiques, université, âgée de 26 à 45 en 1994	29

Liste des tableaux

Tableau 1	Codistribution de la littératie et du revenu	13
Tableau 2	Caractéristiques liées à la forte et à la faible littératie	14
Tableau 3	Régressions du logarithme des résultats de maîtrise des textes schématiques	17
Tableau 4	Régressions de regroupement avec les effets de cohorte	30
Tableau 5	Régressions de regroupement de quantiles avec les effets de cohorte selon les groupes de scolarisation	32
Tableau 6	Régressions avec les variables de la profession et de l'utilisation au travail des compétences en littératie	34
Tableau 7	Régressions des gains	39
Tableau 8	Régressions des gains par quantiles	41
Tableau A.1	Comparaisons de catégorie d'éducation : EIAA et EIACA	46
Tableau A.2	Comparaisons de catégorie d'éducation : EPA	46
Tableau B.1	Régressions du logarithme de la littératie	47

1. Introduction

La littératie des adultes est foncièrement et utilitairement importante. Sen (1999) affirme que nous devrions aspirer à une société où chacun est capable de cultiver ce qu'il perçoit raisonnablement comme étant le bien. Pour ce faire, on doit disposer à la fois de ressources minimales et de caractéristiques que cet auteur appelle « fonctionnements »; la littératie compte parmi les premiers de ces fonctionnements. Si on n'a pas des compétences en littératie, on ne peut participer intégralement et en toute égalité au discours social et politique : sans les instruments de base nécessaires à la réalisation de ses objectifs, on déchoit de son rang de membre égal dans la société (Sen, 1999). Dans tout mouvement d'édification d'une société meilleure, la répartition et la création de compétences en littératie sont d'un intérêt fondamental. La littératie peut aussi être importante utilitairement, puisque celui ou celle qui jouit de meilleures compétences dans ce domaine peut sans doute s'attendre à exploiter de meilleures possibilités d'emploi, de rémunération et donc de bien-être. D'un point de vue sociétal, une main-d'œuvre ayant plus de compétence en littératie sera aussi mieux placée pour s'adapter à l'évolution et faire siennes de nouvelles technologies. Ainsi, l'amélioration de la littératie de l'individu peut avoir des retombées sur la productivité de l'ensemble de l'économie.

Nous examinerons d'abord la répartition des compétences en littératie dans l'économie canadienne, ainsi que leur formation. Dans une large mesure, cette acquisition de compétences doit passer par la scolarisation et l'éducation au foyer. Nous nous attacherons à ces questions, mais pas aussi exhaustivement que ne le permettrait une base de données longitudinales comportant des questions sur la littératie comme l'ELNEJ. Nous regarderons également la nature de la formation de compétences dans les années professionnelles qui suivent la scolarisation. Après avoir tracé les faits saillants du tableau qui peut être dressé de la littératie dans l'économie, nous verrons quelle est l'incidence d'une amélioration des compétences dans ce domaine sur les gains individuels. Il sera question à la fois de l'action causale de la littératie sur les gains et de la corépartition de cette même littératie et du revenu, cette distribution devant nous livrer une mesure plus complète de la qualité du fonctionnement de l'individu dans la société selon Sen. Ainsi, nous examinerons le double rôle fondamental et utilitaire de la littératie dans la société canadienne.

La clé de notre étude, ce sont, bien sûr, les données. Nous utilisons deux versions d'un fonds d'information remarquable où sont combinés les réponses à des questions de démographie et de marché du travail et les résultats des examens en littératie. Notre principale source d'information est le volet canadien de l'Enquête internationale sur l'alphabétisation et les compétences des adultes (EIACA) de 2003. C'est là une très vaste enquête auprès de plus de 22 000 personnes avec suréchantillonnage de la population d'ascendance autochtone. Dans une autre étude, nous avons considéré ce dernier groupe en détail, et nous l'excluons donc de la présente analyse, tout comme la population immigrante. Nous nous servons aussi du volet canadien de l'Enquête internationale sur l'alphabétisation des adultes (EIAA) de 1994 pour dégager un tableau plus complet de l'évolution de la littératie selon l'âge et les cohortes de naissances. Pour prévenir toute confusion, nous parlerons de l'EIACA 2003 et de l'EIAA 1994.

Notre investigation de la littératie à l'aide de l'EIACA 2003 donne de solides résultats. La littératie s'améliore nettement (mais à un taux décroissant) avec le nombre d'années de scolarité. Les niveaux d'éducation parentale sont aussi d'une nette incidence positive sur la littératie, notamment l'éducation maternelle. Par ailleurs, la profession des parents n'a aucune incidence économiquement appréciable ni statistiquement significative sur la littératie une fois que l'on rend compte de l'éducation parentale. De plus, que la mère de l'enquêté ait travaillé lorsque celuici avait 16 ans n'influe pas sur son degré de littératie. Le plus intéressant peut-être, c'est que nous ne relevons guère de liens entre l'âge ou l'expérience professionnelle, d'une part, et la littératie, d'autre part. Nous retrouvons ces mêmes solides résultats dans des travaux antérieurs avec les données de l'EIAA 1994 (Green et Riddell, 2003). De prime abord, l'implication serait que les gens améliorent leur littératie avec la scolarisation et l'éducation au foyer et que les compétences alors acquises ne changent pas pour l'essentiel après les études. Lorsque nous combinons les données de l'EIAA 1994 et de l'EIACA 2003, nous constatons cependant que l'aplatissement de la courbe de compétence en littératie selon l'âge dans les données transversales s'explique en réalité par un ensemble d'effets de vieillissement et de cohorte qui se neutralisent. Notons en particulier que les membres d'une cohorte de naissances perdent en fait de leurs compétences en littératie après avoir quitté l'école. On est aussi nettement porté à croire que les cohortes plus récentes ont moins de compétence en littératie. Cette constatation, qui vaut particulièrement pour les gens mieux scolarisés, ressort surtout de la queue plus mince à droite de la distribution de la littératie (il y a moins de gens qui atteignent de hauts niveaux de littératie). Ainsi, une personne de 35 ans dans l'EIACA 2003 présente en gros les mêmes résultats moyens de littératie qu'une personne de 25 ans dans la même enquête, non pas parce que celle-ci devrait se situer au même niveau dans 10 ans, mais parce que la personne de 35 ans se trouvait à un niveau plus élevé à l'âge de 25 ans (elle appartient à une cohorte ayant un plus niveau de littératie) et a perdu de ses compétences après avoir quitté l'école. Selon ces résultats, les compétences en littératie diminueraient au fil des ans et nous réussirions de moins en moins dans nos tâches d'éducation des générations qui se succèdent. En tenant compte de l'usage qui se fait des compétences en lecture, en écriture et en calcul dans le monde du travail, nous nous demandons en outre si les compétences se perdent à force de ne pas être employées.

2. Données

Nos données viennent de l'Enquête internationale sur l'alphabétisation et les compétences des adultes (EIACA 2003). Il s'agit d'une enquête fascinante qui a eu lieu dans un certain nombre de pays en 2003[1]. Nous exploitons aussi les données de l'Enquête internationale sur l'alphabétisation des adultes (EIAA) de 1994, laquelle a eu lieu au Canada cette année-là. L'EIACA 2003 comporte des questions types sur la démographie, le marché du travail et la rémunération, mais tente aussi de mesurer la littératie et les compétences cognitives qui se forment dans quatre domaines, ceux de la maîtrise des textes suivis et schématiques ou littératie, de la numératie et de la résolution de problèmes (ce dernier domaine n'est pas compris dans l'EIAA 1994). Le plus important peut-être pour notre propos, c'est que l'EIACA 2003 et l'EIAA 1994 ont voulu appréhender non seulement les compétences en numératie et en littératie, mais aussi la capacité d'appliquer les compétences acquises aux situations de la vie quotidienne. Par les questions de ces enquêtes sur la maîtrise des textes suivis, on jauge des capacités diverses, que l'on ait à lire le mode d'emploi sur un flacon d'aspirines ou à « utiliser une annonce d'un département de ressources humaines pour répondre à une question qui emploie d'autres tournures de phrases que celles formulées dans le texte ». Dans les questions sur la maîtrise des textes schématiques par lesquelles on regarde la capacité de repérer et d'utiliser de l'information sous diverses formes, on peut demander d'indiquer des pourcentages par catégories dans un graphique ou d'établir un prix moyen en combinant des éléments d'information. Dans les questions sur la numératie, on peut tour à tour ajouter des éléments d'information sur une formule de commande ou calculer d'après un tableau le pourcentage de calories que donne le gras d'un Big Mac. Il convient de noter pour notre propos que le volet de la numératie a beaucoup changé entre 1994 et 2003 en partie parce qu'on craignait, à la façon dont étaient formulées les questions en 1994, que les compétences en calcul ne puissent être séparées des compétences en lecture. En revanche, les questions sur la maîtrise des textes schématiques et suivis se recouvrent amplement entre ces deux années d'enquête. Les questions sont les mêmes dans une proportion approximative de 45 %. Statistique Canada a aussi renormalisé les résultats de la proportion restante de 55 % en 2003 pour qu'ils soient alignés avec les résultats de 1994. De plus, ce même type d'alignement a eu lieu pour les questions qui étaient identiques dans les deux années. Dans l'analyse qui suit, nous considérons que les résultats en littératie (textes suivis et schématiques) sont parfaitement comparables entre les deux années d'enquête. Nous examinons ensuite cette hypothèse et les données à l'appui.

L'échantillon de l'EIAA 1994 livre des observations sur 5 660 personnes et celui de l'EIACA 2003 renseigne sur beaucoup plus de personnes, soit 23 038. Notre but est de regarder avant tout la formation de compétences en littératie dans l'économie canadienne, d'où la nécessité d'exclure de ces deux échantillons toute personne née à l'étranger, notre point de mire étant le système éducatif canadien. Nous retranchons également les observations du suréchantillon des Premières nations de l'enquête de 2003, réservant un traitement plus direct de cette population pour une autre étude. Les enquêtes portent sur la population de plus de 16 ans, mais nous écartons les gens qui déclarent les études comme activité principale, puisque nous voulons mettre en lumière l'incidence de la scolarisation achevée et l'évolution ultérieure des compétences

ainsi formées. Nos échantillons sont donc respectivement de 3 964 et 14 666 personnes pour l'EIAA 1994 et l'EIACA 2003. C'est la base de notre analyse initiale de la répartition des compétences en littératie dans la société canadienne. Lorsque nous examinerons la création de compétences après les études et l'incidence de la littératie sur les gains, nous nous limiterons cependant à un échantillon ne comprenant pas les travailleurs indépendants ni les salariés dont les gains hebdomadaires sont de moins de 50 $ ou de plus de 20 000 $. Avec cette dernière restriction, nous nous trouvons à éliminer les retraités, les chômeurs et les inactifs. Nous éliminons aussi un petit nombre de travailleurs dont les gains sont des valeurs nettement aberrantes par rapport à la rémunération du reste de l'échantillon. Nous retranchons les travailleurs indépendants, parce que nous désirons examiner la rétribution des compétences sur le marché du travail et que, dans le cas du travail indépendant, il y a rétribution du travail et rétribution du capital. Nous conservons les deux sexes tout au long de notre analyse en faisant la distinction entre hommes et femmes dans certains cas. Enfin, nous mentionnons la pondération d'échantillon fournie avec les données dans l'ensemble de nos tableaux et de nos estimations.

Dans l'analyse des gains, nous nous reportons à une variable « construite » de la rémunération hebdomadaire. Dans l'EIACA 2003, on s'enquiert d'abord auprès des gens de leur période de paie normale et ensuite de leurs gains normaux dans cette période. À l'aide des réponses, nous élaborons une variable individuelle de la rémunération hebdomadaire. Si la période de paie indiquée est mensuelle, nous divisons les gains déclarés par 4 333.

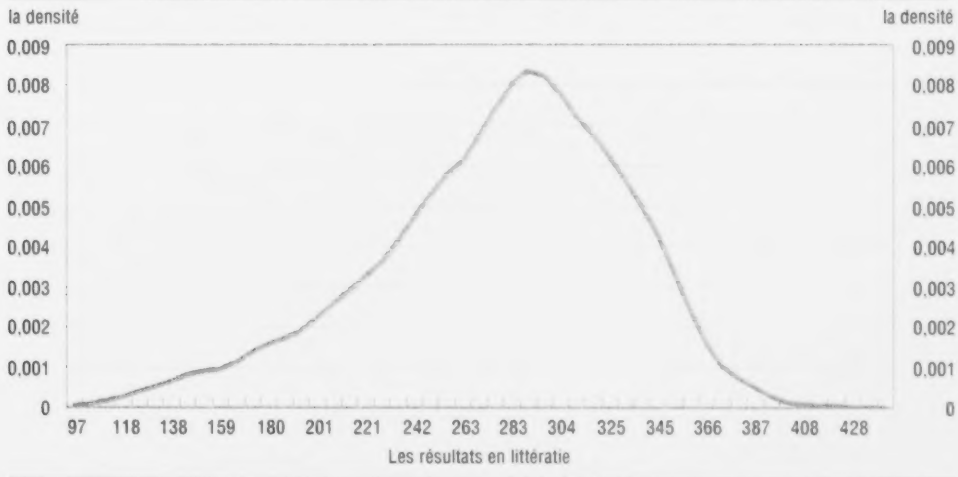
Un trait distinctif de ces données est l'étroite corrélation qui existe entre les divers résultats de littératie. La corrélation entre les résultats de maîtrise des textes suivis et schématiques est de 0,96 dans l'EIACA 2003; elle est de 0,92 entre le domaine des textes schématiques et la numératie et de 0,92 entre ce même domaine et la résolution de problèmes. Qui plus est, une analyse en composantes principales dégage pour l'essentiel une composante en équipondération pour les quatre grands résultats. Ainsi, une simple moyenne des résultats nous livre le gros de l'information des quatre. Comme les résultats sont à ce point en corrélation, nous en prendrons un en particulier au moment de décrire la répartition et la formation des compétences en littératie. À notre avis, il vaut mieux s'attacher à une seule variable significative qu'à un mélange des résultats dans cette partie de l'analyse, parce qu'on distingue plus nettement comment les résultats se forment. Nous regarderons plus précisément les résultats de maîtrise des textes schématiques. Nous avons le choix entre textes schématiques et textes suivis, les résultats étant communs à l'EIAA 1994 et à l'EIACA 2003. De plus, les résultats pour le domaine des textes schématiques semblent le plus en saturation dans l'analyse factorielle. Aucunes des conclusions que nous offrons ne changerait si nous faisions intervenir les résultats relatifs aux autres domaines. Dans l'analyse des gains sous l'angle des effets généraux de la littératie, nous nous reportons à une simple moyenne des quatre grands résultats, ce qui permet aussi une comparaison plus claire avec nos travaux antérieurs.

3. Répartition de la littératie

Nous commençons par une simple description de la répartition des compétences dans le domaine des textes schématiques en nous fondant sur les données de l'EIACA 2003. Au graphique 1, nous présentons une estimation de densité noyau de ces compétences pour notre échantillon. La courbe de densité est légèrement en asymétrie négative avec une moyenne de 280, une médiane de 287 et un écart-type de 50. Qu'il y ait asymétrie négative dans ce qui serait normalement une distribution en asymétrie positive comme la distribution lognormale tient peut-être à l'insuffisance des questions pouvant permettre de différencier les répondants ayant un haut niveau de littératie. Pour l'échantillon, le résultat minimal est de 84 (niveau extrêmement bas de maîtrise des textes schématiques) et le résultat maximal, de 436[2]. De l'inégalité de cette distribution témoignent un coefficient de Gini de 0,107 et une valeur de 0,513 pour le logarithme du rapport 90[e] percentile au 10[e] percentile. Mettons le tout en perspective en disant que, pour le revenu familial avant impôt et transferts au Recensement canadien de 2001, le coefficient de Gini est de 0,438 et le logarithme du ratio du 90 à 10, de 3,48 (Frenette et coll., 2006). Ainsi, la répartition de la littératie est bien moins inégale que celle du revenu du travail, ce dont on ne s'étonnera pas puisque les compétences en littératie ne sont qu'un des facteurs de formation du revenu du travail.

Graphique 1

Les résultats à l'échelle des textes schématiques, 2003



L'autre comparaison possible avec ces données peut porter sur le degré d'inégalité de la répartition de la littératie en 1994. Dans ce cas, le coefficient de Gini est de 0,151 et le logarithme du ratio du 90 à 10, de 0,76. Ainsi, la distribution se fait bien moins inégale dans les neuf ans compris entre les deux années d'enquête. C'est en partie à cause d'une amélioration à l'extrémité inférieure de la distribution de la littératie. En 1994, le 10e percentile est de 160 et, en 2003, de 197. La nouvelle n'est pas si bonne cependant si on considère que, de 1994 à 2003, le 95e percentile est en fait tombé de 359 à 351. Ainsi, cette baisse de l'inégalité dans l'échelle de répartition de la littératie s'explique à la fois par une amélioration appréciable à l'extrémité inférieure de la distribution et par une certaine détérioration à l'extrémité supérieure. Nous reviendrons sur ces différences, mais pour le moment nous chercherons à caractériser la répartition de 2003.

Au tableau 1, nous dépeignons une corépartition de la littératie et du revenu familial et indiquons le pourcentage d'observations dans chaque cellule d'une grille définie par les cinq quintiles de la distribution des compétences en textes schématiques et les cinq quintiles de la distribution du revenu du ménage (avant impôt et transferts). Entre le revenu du ménage et la littératie, la corrélation est de 0,31, indice d'une association positive, mais plutôt lâche, ce qui, au tableau 1, se remarque aux valeurs de la diagonale principale (aux traits plus prononcés) qui ne sont pas beaucoup plus élevées que les autres valeurs. La cellule la plus intéressante est peut-être celle qui correspond au quintile du bas des deux distributions et qui contient un peu plus de 8 % de la population. On peut considérer que les intéressés sont victimes d'une double pauvreté en compétence et en revenu. Dans cette cellule, tout le monde a un bas niveau de littératie pour reprendre l'expression de Crompton (1996). Pour ce qui est de la maîtrise des textes schématiques, cet auteur décrit les « marginaux de la littératie » comme étant sans doute incapables de relever sur un horaire d'autobus quand le dernier autobus quittera tel ou tel arrêt le samedi soir (Crompton, 1996). Ce sont des gens lourdement handicapés dans leur fonctionnement en milieu social. On ne s'étonnera sans doute pas que ces doubles victimes soient extrêmement peu scolarisées et que 80 % d'entre elles aient décroché de l'école secondaire. Il en va de même de leurs antécédents familiaux, car plus de 90 % d'entre elles ont aussi des pères décrocheurs. En revanche, les 7,4 % de gens qui appartiennent aux deux quintiles supérieurs de la littératie et du revenu sont très scolarisés et appartiennent aux milieux instruits. Une proportion de 30 % des intéressés sont des diplômés d'université et une proportion supplémentaire de 21 % ont fait des études universitaires supérieures. Le quart des pères sont aussi allés à l'université, proportion bien supérieure à celle que l'on peut relever dans la population en général où un peu plus de 9 % des pères sont titulaires d'un grade universitaire.

Tableau 1
Codistribution de la littératie et du revenu

	Quintile de revenu				
Quintile de littératie	1er quintile	2e quintile	3e quintile	4e quintile	5e quintile
	pourcentage				
1er quintile	8,3	5,6	3,4	1,6	0,8
2e quintile	4,4	5,3	4,4	3,4	2,4
3e quintile	2,5	3,9	4,8	4,9	4
4e quintile	1,9	3,2	4,4	5,2	5,4
5e quintile	1,3	2,4	4,2	4,9	7,4

Nota : Tableau fondé sur des calculs à l'aide des données de l'EIACA 2003. La littératie correspond à la maîtrise des textes schématiques. Le revenu est le revenu du ménage avant impôt et transferts. Les valeurs des cellules correspondent aux pourcentages respectifs de l'ensemble des observations selon les quintiles des distributions de la littératie et du revenu. L'addition des chiffres aux diverses lignes et colonnes peut ne pas donner 20 à cause des arrondis.

Pour répandre un autre éclairage sur les difficultés qui peuvent être liées à la faiblesse de la littératie, nous livrons au tableau 2 des mesures de la participation sociale et du bien-être individuel des gens ayant un faible niveau de littératie et un haut niveau de littératie. Les gens avec un faible niveau de littératie sont ceux qui se situent au niveau 1 parmi les cinq niveaux de maîtrise des textes schématiques. C'est là un groupe extrêmement défavorisé et sans doute incapable d'accomplir des tâches comme reconnaître des similarités entre plusieurs tableaux (Crompton, 1996). Les gens avec un haut niveau de littératie sont ceux qui appartiennent aux deux niveaux supérieurs de maîtrise. Les premiers seraient plus susceptibles d'avoir voté aux élections municipales que les seconds, mais c'est la seule mesure de participation où ils dominent. Ceux avec un haut niveau de littératie sont plus susceptibles d'avoir participé aux élections fédérales ou provinciales, mais la différence n'est pas marquée. Les divergences sont importantes cependant dans le cas de l'appartenance aux organismes politiques et aux groupements communautaires et scolaires, ceux avec un haut niveau de littératie ayant au moins deux fois plus de chances de participer aux activités de tels groupes. Nous ne prétendons pas qu'il y ait rapport de causalité entre la littératie et la participation (et qu'en augmentant la littératie, on augmenterait la participation), mais le constat de franche corrélation fait bel et bien ressortir l'importance de la littératie pour une pleine participation à la vie sociale au dire même de Sen (1999). Ceci ce manifeste dans le fait que près de 11 % des gens ayant un faible niveau de littératie aient évoqué l'anxiété qu'ils éprouvent à faire de simples calculs comme le calcul d'un pourboire. Là où le lien se fait le plus étroit, c'est pour l'autoévaluation de la santé. Environ 76 % des gens avec un haut niveau de littératie se disaient en excellente ou en très bonne santé comparativement à 30 % seulement des gens avec un faible niveau de littératie. Pour nous assurer que ce n'est pas seulement le mauvais état de santé qui fait que les gens ont de la difficulté à réussir aux examens de littératie, nous calculons ces pourcentages après avoir écarté les gens dont nous savions qu'ils n'avaient pas répondu au questionnaire de base à cause d'ennuis de santé ou d'une incapacité. L'écart santé-littératie tient en partie à ce que les gens avec un faible niveau de littératie soient bien plus âgés, ce qui révèle l'existence d'une variation intergénérationnelle de la littératie. Toutefois, si nous excluons les gens de plus de 65 ans et estimons la régression probabiliste linéaire d'une variable fictive égale à l'unité pour l'autodéclaration d'une très bonne ou excellente santé en fonction de l'âge, des années de scolarité et du revenu familial, ainsi que d'une variable fictive correspondant aux gens avec un faible niveau de littératie, cette dernière variable garde un coefficient de -0,16 et est hautement significative statistiquement (avec une erreur-type de 0,041). Ainsi, les gens avec un faible niveau de littératie ont environ 16 % moins de chances de se dire en bonne santé même après prise en compte de l'âge, des années de scolarité et du revenu familial. Là encore, on ne peut clairement parler de rapport de causalité, mais on peut sérieusement s'interroger sur la façon dont peuvent fonctionner les gens avec un faible niveau de littératie dans la société canadienne.

Tableau 2

Caractéristiques liées à la forte et à la faible littératie

Variable	Groupe de faible littératie	Groupe de forte littératie
	pourcentage	
Participation aux élections municipales	67,9	61,7
Participation aux élections fédérales	79,5	84,2
Appartenance à des organismes politiques	2,6	6,7
Appartenance à des organismes communautaires ou scolaires	11,4	47,4
Satisfaction à l'égard de la vie	78,3	82,5
Bonne santé	29,9	75,7
Anxiété à propos des chiffres	10,7	3,8
Revenu du ménage (médian)	30 000	90 000

Nota : Calculs à l'aide des données de l'EIACA 2003. Toutes les valeurs sont en pourcentage sauf le revenu du ménage exprimé en dollars. Le groupe de faible littératie comprend les gens qui se situent au niveau 1 pour les cinq critères de vérification de la maîtrise des textes schématiques. Le groupe de forte littératie comprend les gens qui se situent au niveau 4 ou 5 pour ces mêmes critères.

4. Formation des compétences en littératie

4.1 Incidence de la scolarisation et des antécédents

Nous nous reporterons aux données tant de l'EIAA 1994 que de l'EIACA 2003 pour examiner les sources de littératie. Nous effectuerons une suite de régressions où, une fois de plus, les résultats de maîtrise des textes schématiques seront notre mesure représentative. Plus précisément, nous prendrons le logarithme de ces résultats comme variable dépendante de manière à pouvoir interpréter les coefficients que nous estimerons comme indiquant les conséquences sous forme de variations en pourcentage de la littératie. Nous commencerons par le fonds d'information bien plus vaste de l'EIACA 2003 et cernerons l'incidence de l'éducation au foyer et de la scolarisation sur les résultats individuels en littératie.

Avant de présenter nos estimations, nous décrirons brièvement un modèle heuristique de la formation des compétences en littératie. Grâce à ce modèle, nous pourrons mieux mettre nos estimations en contexte, ainsi que mieux orienter nos réflexions sur les questions d'identification. Considérons un simple modèle où, à la naissance, les gens reçoivent une double caractérisation par leur capacité propre et les ressources parentales. Par ressources parentales, nous entendons quelque chose de plutôt général, c'est-à-dire à la fois le revenu des parents et le fait que ceux-ci soient à la fois désireux et capables de favoriser l'éducation et la littératie de leurs enfants. À l'âge préscolaire, les enfants commencent à développer des compétences en littératie à l'aide de ses caractéristiques fondamentales (la capacité propre et les ressources parentales). Une fois qu'ils entrent dans le système scolaire, ces mêmes caractéristiques seront en interaction avec des caractéristiques de l'école comme la qualité pédagogique, la taille des classes et les attitudes et aptitudes des pairs. Les nouveaux acquis en littératie avec chaque année de scolarité sont alors fonction de la capacité propre, des ressources parentales, des caractéristiques scolaires et du degré de littératie en début de période. L'interaction de ces facteurs peut s'avérer complexe. Il y a accumulation de gains de littératie jusqu'à l'âge de déscolarisation, après quoi les élèves décident chaque année de poursuivre leurs études jusqu'à la fin de leur fréquentation du palier secondaire. Une telle décision sera fonction une fois de plus de la capacité propre, des ressources parentales et des caractéristiques scolaires, mais aussi sans doute des compétences en littératie qui se sont formées jusque-là. Plus l'élève a des compétences en littératie, moins il trouve pénible de se trouver à l'école et plus il est enclin à décider d'y passer une année de plus. Après les études secondaires, la poursuite de la scolarisation sera déterminée à la fois par une décision propre de continuer à étudier et une décision d'admission du collège ou de l'université. Cette dernière décision sera sans doute fonction de la littératie de l'élève selon les notes qu'il a obtenues. Ainsi, scolarisation et littératie sont codéterminées et un surcroît de scolarisation mène à un surcroît de littératie et vice versa, plus particulièrement après l'âge de déscolarisation. Si nous tenons aussi compte des attentes, le rapport entre scolarisation et littératie peut se révéler encore plus étroit. Les gens qui ne s'attendent pas à continuer à étudier après l'âge de déscolarisation pourraient en effet rationnellement s'aviser de sous-investir dans l'acquisition de compétences en littératie pendant leur période de scolarisation.

La formation de compétences sera probablement plus difficile après le départ de l'école. Elle peut se faire au travail si l'intéressé a besoin de compétences pour l'accomplissement de certaines tâches, mais dans les autres cas il lui faudra investir activement dans ses heures de loisirs. En fait, il paraît fort possible que les gens perdent de leurs compétences en littératie après les études s'il s'agit de compétences qui se dégradent à force de ne pas être employées.

Nous voulons caractériser le plus grand nombre possible de facteurs de formation de compétences en littératie. Nous nous intéressons en particulier aux liens entre la littératie et les ressources parentales, car l'équité est foncièrement en cause. Dans la mesure même où la littératie d'une génération dépend des ressources de la génération qui la précède, on pourrait juger que les différences de littératie tiennent à des caractéristiques indépendantes de la volonté des intéressés. C'est là l'argument par excellence pour une politique de redistribution aux yeux de philosophes comme Dworkin et Rawls. Nous nous intéressons aussi aux liens entre la scolarisation et la littératie, puisque c'est là une grande voie par laquelle nous pouvons espérer agir sur la répartition des compétences en littératie. Il y a enfin la question de savoir si les compétences s'appauvrissent ou s'enrichissent après l'école et en quoi ce processus a à voir avec les caractéristiques de l'emploi de chacun. Si la littératie se perd à force de ne pas être « employée », il y a peut-être lieu d'adopter des politiques où, par exemple, on subventionnera les entreprises pour des activités de maintien des compétences (possibilités de retour aux études pour les salariés, etc.). Un grand nombre des paramètres d'intérêt font intervenir des rapports de causalité difficiles à établir en définitive. Nous nous efforcerons d'estimer les paramètres causaux si les données le permettent, mais le gros de ce qui suivra se présente nécessairement sous la forme de liens de corrélation, et non pas de liens nets de causalité.

À la colonne MCO 1 du tableau 3, nous présentons notre régression la plus simple par les moindres carrés ordinaires (MCO) où la variable dépendante est le logarithme des résultats individuels de maîtrise des textes schématiques et les variables indépendantes sont l'âge, l'âge en valeur quadratique, les années de scolarité, les années de scolarité en valeur quadratique et une variable fictive du sexe. Grâce à la taille appréciable de notre échantillon, toutes les variables sont significativement différentes de zéro au niveau de 1 % ou de 5 %, mais ce n'est pas dire pour autant que l'incidence effective est appréciable. Ainsi, nos estimations disent que les femmes ont moins de compétence en littératie que les hommes (ce qui dépend de la scolarisation et de l'âge), mais seulement dans une proportion de 1,4 %. De même, les coefficients de l'âge et de l'âge au carré sont statistiquement très significatifs, mais ils impliquent ensemble que l'incidence d'une année de scolarité de plus sur la littératie est en réalité de -0,1 % à l'âge de 30 ans[3]. Qu'il n'y ait essentiellement aucun rapport entre la littératie et l'âge ou l'expérience professionnelle tient une place de choix dans le traitement de la question par Green et Riddell (2003). La seule relation qui soit économiquement appréciable est celle de la littératie et de la scolarisation. Une année de scolarité de plus lorsque l'intéressé en compte déjà 12 accroît la littératie de 3,2 %. Cela ressemble à ce que Green et Riddell (2003) ont calculé à l'aide des données de l'EIAA 1994[4].

À la colonne MCO 2 du tableau 3, nous ajoutons les variables de la scolarisation parentale et de la situation d'immigrant à notre régression MCO. L'introduction de ces variables n'a pour ainsi dire aucun effet sur les variables du sexe et de la scolarisation, mais leur inclusion fait presque doubler le coefficient de l'âge. Comme le coefficient de la variable de l'âge en valeur quadratique devient aussi plus négatif, l'incidence nette de l'âge est toujours très restreinte. Les variables de la scolarisation parentale sont ensemble très différentes significativement de zéro, mais ce qui étonnera peut-être, c'est que l'effet se situe presque entièrement aux niveaux inférieurs de la scolarisation des parents. Que la mère ou le père ait décroché de l'école secondaire diminue la littératie moyenne de 3 % à 4 %. Il reste que la scolarisation parentale au-delà des études secondaires n'influe plus sur la littératie. Aspect intéressant, que l'on ignore le niveau de scolarité de la mère ou du père (tel est le cas pour environ 8 % de l'échantillon) a en soi un effet marqué,

puisque cette indétermination diminue la littératie d'environ 6 %. Nous n'avions inclus cette variable que pour pouvoir conserver les observations pour lesquelles l'indication de la scolarisation parentale manquait, mais il paraît possible que cette catégorie d'ignorance permette d'appréhender quelque chose de bien réel. Par exemple, il se peut que les enfants qui ignorent la scolarité de leur mère ou de leur père n'avaient pas de rapports étroits avec l'intéressé. Ainsi, le coefficient estimé pourrait rendre compte du degré de formation des compétences en littératie par participation directe des parents. Enfin, que le père soit issu de l'immigration est en association modérée avec la littératie (augmentation de 1,5 %), mais que la mère vienne de l'étranger n'a aucune incidence. Nous avons aussi analysé des spécifications avec un ensemble de variables fictives de professions des parents, mais ensemble ces variables n'étaient jamais statistiquement significatives. Notons en particulier qu'une vérification de l'hypothèse selon laquelle l'ensemble de variables fictives de la profession du père est d'une incidence nulle dégage une valeur p de 0,13. Pour la profession de la mère, la valeur p est de 0,79. Nous constatons aussi qu'une variable fictive pour le travail de la mère lorsque l'enfant avait 16 ans n'a aucun effet statistiquement significatif. Dans l'ensemble, ces résultats font voir un lien étonnamment faible entre la littératie et les caractéristiques parentales. Il n'y a que la scolarisation qui aurait un effet appréciable sur la formation des compétences en littératie.

Tableau 3

Régressions du logarithme des résultats de maîtrise des textes schématiques

Variable	MCO 1	Erreur-type	MCO 2	Erreur-type	MCO 3	Erreur-type	MCO 4	Erreur-type	IV	Erreur-type
Sexe féminin	-0.014**	(0.0046)	-0.012**	(0.0045)	-0.0083*	(0.0045)	-0.012**	(0.0044)	-0.016**	(0.0051)
Années de scolarité	0.058***	(0.0053)	0.054***	(0.0049)	0.053***	(0.0048)	0.024***	(0.0008)	0.052***	(0.013)
Années de scolarité en valeur quadratique	-0.0011***	(0.0002)	-0.0011***	(0.0002)	-0.0011***	(0.0002)	...	...	...	...
Âge	0.0035***	(0.0008)	0.0064***	(0.0008)	0.0066***	(0.0008)	-0.0068***	(0.0008)	-0.0003	(0.0036)
Âge en valeur quadratique	-0.0076***	(0.0009)	-0.0099***	(0.0009)	-0.011***	(0.0009)	-0.011***	(0.0009)	-0.0029	(0.0039)
Scolarisation de la mère										
Moins que les études secondaires	...	...	-0.037***	(0.0059)	-0.038***	(0.0058)	-0.028***	(0.0058)	0.0007	(0.016)
Études postsecondaires incomplètes	...	...	-0.0077	(0.0067)	-0.0084	(0.0066)	-0.0046	(0.0065)	-0.020*	(0.011)
Baccalauréat ou plus	...	...	0.0094	(0.0102)	0.0075	(0.0098)	0.0076	(0.01)	-0.0050	(0.016)
Non-déclaration	...	...	-0.067***	(0.0123)	-0.069***	(0.0122)	-0.048***	(0.011)	-0.0047	(0.024)
Scolarisation du père										
Moins que les études secondaires	...	...	-0.032***	(0.0067)	-0.03***	(0.0067)	-0.029	(0.0067)	-0.008	(0.012)
Études postsecondaires incomplètes	...	...	0.0055	(0.0073)	0.006	(0.0071)	0.0062	(0.0073)	-0.008	(0.012)
Baccalauréat ou plus	...	...	0.0126	(0.0083)	0.0154*	(0.0082)	0.011	(0.0082)	-0.025	(0.02)
Non-déclaration	...	...	-0.056***	(0.012)	-0.055***	(0.012)	-0.065***	(0.011)	-0.021	(0.023)
Mère immigrante	...	...	0.0056	(0.0077)	0.0053	(0.0076)	-0.0009	(0.0077)	-0.013	(0.011)
Père immigrant	...	...	0.015**	(0.0071)	0.016**	(0.007)	0.0054	(0.0071)	-0.0084	(0.011)
Obtention de bonnes notes en mathématiques	...	...	...	...	0.027***	(0.0055)	...	...	...	...
Lenteur de l'assimilation des matières scolaires	...	...	...	...	-0.026***	(0.0057)	...	...	...	...
Constante	5.08***	(0.04)	5.09***	(0.038)	5.09***	(0.038)	5.29***	(0.02)	5.04***	(0.12)
Nombre d'observations	14 527		14 527		14 527		13 868		13 868	
R^2	0.49		0.51		0.52		0.47		0.310	

... n'ayant pas lieu de figurer

* Niveaux de signification statistique à 10 %.

** Niveaux de signification statistique à 5 %.

*** Niveaux de signification statistique à 1 %.

Nota : Les valeurs instrumentales de la scolarisation à la dernière colonne sont les variables fictives de la province de fréquentation de l'école secondaire et ces dernières variables sont en interaction avec l'âge.

Comme nous l'avons indiqué, la littératie et les années de scolarisation sont sans doute en codétermination, auquel cas notre coefficient MCO pour la scolarisation nous livre une estimation biaisée de l'incidence de la scolarisation sur la littératie. Nous traitons ce problème de deux façons. D'abord, il peut y avoir biais parce que la corrélation entre littératie et scolarisation tient à une capacité inobservée. Si les gens très capables ne jugent pas particulièrement coûteux d'améliorer leur niveau de littératie ou de se scolariser, nous pourrions dégager un fort coefficient positif pour la scolarisation dans notre régression, parce que les années de scolarité rendent compte de la capacité propre plutôt que de l'action causale de la scolarisation sur la littératie. On peut résoudre le problème par une mesure de cette capacité : une fois que nous prenons la capacité en compte, la relation dégagée entre la scolarisation et la littératie ne peut être due à un terme omis de la capacité propre. À noter toutefois que, dans bien des études où on essaie de prendre en compte la capacité (dans des régressions de gains, par exemple), on se reporte en réalité à des résultats d'examen un peu comme nos examens en littératie. Ce dont nous aurions besoin, ce sont des résultats d'examen à un très jeune âge, avant même que ne s'amorce effectivement le processus que nous voulons ici étudier. Comme nous n'avons pas de tels résultats, nous tenons compte de la capacité propre à l'aide de variables qui y sont vraisemblablement liées. Dans la troisième de nos régressions en particulier, nous incluons une variable fictive qui prend la valeur de 1 en cas d'acceptation simple ou appuyée d'un énoncé selon lequel l'intéressé aurait obtenu de bonnes notes en mathématiques à l'école, et une autre variable fictive où l'unité est l'acceptation correspondante d'un énoncé selon lequel il aurait souvent eu du mal à suivre les professeurs en classe. Ces deux énoncés rendent vraisemblablement compte de la capacité innée. Ce sont deux variables significatives avec une augmentation de 2,7 % de la littératie pour l'obtention de bonnes notes en mathématiques et une diminution de 2,6 % pour la lenteur de l'assimilation de matières scolaires. Il reste que l'inclusion de ces variables n'a presque aucun effet sur les autres coefficients estimés. Le plus remarquable est que cette inclusion ne change en rien l'incidence de la scolarisation sur la littératie.

Une autre façon de résoudre le problème de détermination d'une action causale est de trouver une variable instrumentale pour la scolarisation et d'estimer la régression par la méthode des moindres carrés à deux degrés. Ce sont là des variables qui influent directement sur la variable endogène du côté droit, mais qui n'agissent sur la variable dépendante que par sa relation avec cette même variable endogène. Pour l'essentiel, nous pouvons estimer les effets en fonction de la partie de la variation de la variable endogène qui sera en corrélation avec la variable instrumentale et donc sans corrélation avec le terme d'erreur. Il devient possible d'estimer en toute cohérence l'incidence véritable de la scolarisation sur la littératie si deux conditions sont réunies : 1) la variable instrumentale influe en réalité sur la scolarisation et 2) elle n'appartient pas à la régression qui détermine notre variable dépendante en soi. Nous nous reportons à la province de résidence à la dernière fréquentation du palier secondaire ou primaire, et ce, en pleine interaction avec l'âge que comme variable instrumentale de la scolarisation. L'idée est que les ressources publiques de scolarisation variaient selon les provinces et les générations, d'où des résultats différents de scolarisation pour des gens semblables par ailleurs. Nous ne voyons pas pourquoi la province de résidence influerait sur la littératie autrement que par la variation des résultats de scolarisation. Il importe que, dans ce traitement, nous tenions également compte de la province actuelle de résidence tant au premier degré (régression de la scolarisation) qu'au second (régression de la littératie). La province actuelle de résidence pourrait être liée à la littératie si les gens plus scolarisés choisissent de migrer vers des provinces où les emplois de forte littératie sont proportionnellement plus nombreux et si les gens ayant un faible niveau de littératie optent au contraire pour des provinces où abonderaient les emplois dans les industries de ressources, par exemple. Dans ce cas et dans la mesure même où il y a corrélation des provinces de résidence à l'époque de la fréquentation de l'école secondaire et à l'heure actuelle, cette variable appréhenderait cet effet de migration plutôt que l'effet de scolarisation que nous

visons. En tenant compte de la province actuelle de résidence, le problème est résolu, et nous cernons en réalité l'effet de scolarisation pour des gens qui résident actuellement dans la même province, mais qui ont été scolarisés dans d'autres provinces à des époques différentes.

On trouvera à la colonne IV les résultats de cette estimation à deux degrés par les moindres carrés où la province de résidence à l'époque de la fréquentation de l'école secondaire fait fonction de variable instrumentale. Nous sommes incapables de trouver une variable distincte pour la scolarisation en valeur quadratique. Une façon normale de le faire serait de prendre pour la valeur quadratique des termes d'ordre supérieur de la variable instrumentale, mais comme il s'agit d'une variable fictive (ou plus précisément d'un ensemble de telles variables), les termes d'ordre supérieur sont déjà cette variable fictive et nous nous retrouvons sans variable instrumentale distincte pour la scolarisation au carré. Nous estimons donc une spécification qui comprend seulement la variable linéaire de la scolarisation. À la colonne MCO 4, nous présentons des estimations de régression MCO par la même spécification à des fins de comparaison[5].

Au premier degré (forme réduite) de la régression où le nombre d'années de scolarité est la variable dépendante, les covariables sont toutes celles de notre principale régression de la littératie, à savoir l'âge, l'âge en valeur quadratique, la scolarisation et l'immigration des parents et la province actuelle de résidence. Il y a aussi le jeu de variables fictives qui représentent la province de fréquentation de l'école secondaire. Comme on pouvait s'y attendre, la scolarisation parentale est en corrélation positive étroite avec le nombre d'années de scolarité. Si les parents sont issus de l'immigration, il y a aussi un lien avec la prolongation de la scolarisation. Les variables de la province à l'époque des études secondaires et leurs éléments d'interaction avec l'âge sont ensemble très significatifs statistiquement, indice que la condition de l'importance du lien de détermination entre la variable instrumentale et la variable endogène se trouve respectée[6]. On peut voir les estimations du premier degré à l'annexe B.

Dans les résultats de la colonne MCO 4 les effets estimés de la scolarisation sont hautement significatifs. Leur ordre de grandeur est celui que nous avons indiqué comme l'incidence en pourcentage d'une année de scolarité de plus sur la littératie lorsqu'on compte déjà 12 années de scolarité dans une spécification quadratique. L'estimation VI de l'effet de scolarisation à la colonne IV est environ deux fois plus élevée cependant, ce qui implique l'existence d'effets plus marqués que ceux que nous avons estimés dans les spécifications MCO. Ce qui est intéressant, c'est qu'une fois que nous prenons une variable instrumentale pour la scolarisation, les variables des caractéristiques parentales perdent de l'importance et de leur signification statistique. C'est peut-être dire que les effets de ces caractéristiques que nous avons estimés étaient en réalité des effets de lieu de scolarisation ou, du moins, qu'il y a une forte collinéarité des variables des caractéristiques parentales et du lieu de scolarisation. Comme dans les spécifications antérieures, le sexe et l'âge continuent à être d'une faible incidence sur la littératie. Par souci de concision, nous ne présentons pas les coefficients correspondant aux variables fictives de la province, mais précisons que les provinces de l'Atlantique et l'Ontario ont des niveaux de littératie essentiellement convergents, le Québec, des niveaux significativement inférieurs, et les Prairies et la Colombie-Britannique, des niveaux significativement supérieurs. Notre grande conclusion est que, si les hypothèses de choix de notre variable instrumentale sont justes, la scolarisation a une forte action causale sur la littératie et qu'elle en est la source prédominante[7]. Pour mettre l'effet estimé en perspective, disons que, d'après les estimations IV, une prolongation de quatre ans de la scolarisation (si on passe, par exemple, d'un diplôme du palier secondaire à un grade du palier universitaire) implique une augmentation de 21 % de la littératie, assez pour que l'on passe de la médiane au 90e percentile environ dans l'échelle de répartition de la littératie en 2003.

4.2 Effets d'âge et de génération

Un des résultats les plus frappants de nos régressions initiales est le manque de signification de la variable de l'âge. De prime abord, on pourrait penser que les gens ne perdent ni ne gagnent en littératie après avoir quitté l'école. Les compétences en littératie se formeraient avant tout à l'école et se maintiendraient tout simplement par la suite. Ce constat possible se trouve renforcé si nous remplaçons l'âge et l'âge en valeur quadratique par l'expérience et l'expérience élevée au carré dans la spécification de régression au tableau 3[8]. Les coefficients de la scolarisation, de la scolarisation au carré et de la variable fictive du sexe féminin sont pour ainsi dire identiques à ceux du tableau 3. Le coefficient de la variable de l'expérience est très faible (0,00051) et l'erreur-type s'établit à 0,00045.

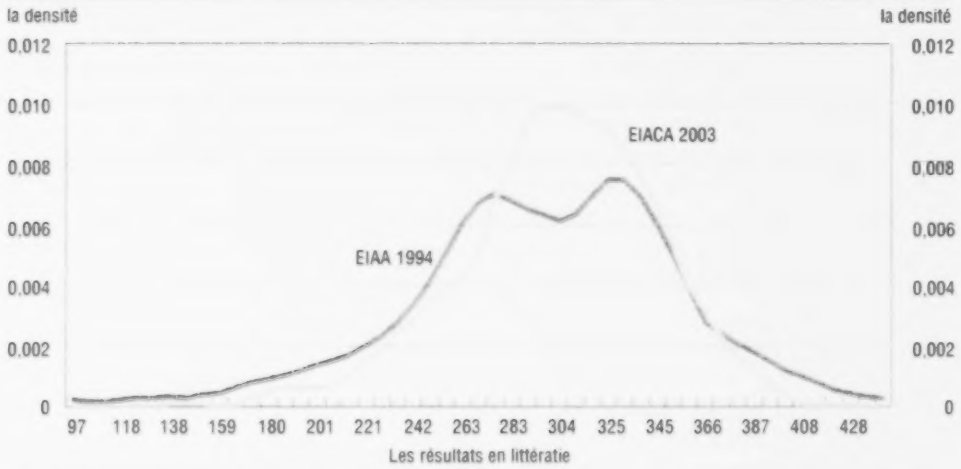
L'interprétation du coefficient de l'âge dans une régression transversale doit se faire avec une certaine prudence cependant. Ainsi qu'en témoigne le vaste examen de la question dans les études consacrées à la rémunération des immigrants, les différences entre deux groupes d'âge dans des données transversales seraient attribuables à des effets véritables d'âge et de génération dans diverses combinaisons possibles. Si nous sommes tentés de voir dans le niveau de littératie des 35 ans dans l'EIACA 2003 un reflet du niveau de littératie des 25 ans dans un horizon de 10 ans, nous devons garder à l'esprit que les 35 ans appartiennent à une génération plus âgée et que leur littératie observée est à mettre au compte d'une combinaison de différences générationnelles et d'effets de vieillissement. Ce n'est que lorsqu'il n'y a pas de différences systématiques entre les générations que le profil transversal littératie–âge rend véritablement compte de l'incidence du vieillissement sur la littératie.

Pour une étude plus complète des effets de génération (souvent appelés effets de cohorte) et d'âge, il faut recourir soit à d'authentiques données longitudinales, soit à au moins deux ensembles de données transversales permettant de suivre des cohortes « synthétiques » dans le temps. Nous pouvons employer à cette fin l'EIAA 1994 et l'EIACA 2003. Plus précisément, nous pouvons relever, dans la version à grande diffusion des données de l'EIAA 1994, une suite de tranches quinquennales d'âge des répondants (26 à 35, 36 à 45, 46 à 55, 56 à 65 et 65 ans et plus). Comme la variable de l'âge est continue dans l'EIACA 2003, nous pouvons élaborer des tranches d'âge qui correspondent aux tranches d'âge de l'EIAA 1994 après décalage de 9 ans (35 à 44, 45 à 54, 55 à 64, 65 à 74 et 75 ans et plus). Notre cohorte 1 sera formée des gens âgés de 26 à 35 ans en 1994 et de 35 à 44 ans en 2003. Nous numéroterons les autres cohortes par ordre croissant. Si ces deux enquêtes comportent des échantillons indépendants, chacune livre une estimation sans biais de la répartition de littératie d'une cohorte à deux moments, et il est possible de suivre les progrès de cette cohorte dans le temps. Dans le cadre de cet examen, nous décomposerons les cohortes selon la scolarisation. Le couplage des années d'enquête ne donnerait pas d'estimation sans biais des progrès des groupes cohorte–scolarisation dans ce cas si la composition en scolarisation de ces groupes varie dans le temps. Si une partie des diplômés de l'école secondaire au départ ont poursuivi leurs études, il est impossible d'affirmer que la littératie moyenne des diplômés de l'école secondaire de 35 à 44 ans dans l'EIACA 2003 est une estimation fidèle de l'évolution des diplômés de l'école secondaire de 26 à 35 ans que nous observons dans l'EIAA 1994 sur une période de 9 ans. Dans ce cas, la littératie moyenne en 2003 combinerait les effets de vieillissement et une désélection automatique dans le groupe des diplômés de l'école secondaire. Pour éviter la difficulté, nous nous attacherons aux gens de plus de 25 ans, stade après lequel les variations ponctuelles de scolarisation sont rares parmi nos grandes catégories. Il convient en outre de noter que la cohorte la plus âgée (gens de plus de 65 ans en 1994) est particulière, étant d'une composition susceptible de varier dans le temps à cause des décès. Nous l'incluons dans notre analyse, mais nous n'insisterons pas outre mesure sur ses résultats dans nos conclusions.

Nous commençons par des estimations de densité des résultats de maîtrise des textes schématiques dans l'EIAA 1994 et l'EIACA 2003 en ventilation par l'âge, la scolarisation et la cohorte. Au graphique 2, nous livrons des estimations de densité des gens âgés de 26 à 35 ans en 1994 et des gens du même âge en 2003. Ces valeurs de densité de littératie sont celles de la cohorte la plus jeune en 2003 et de la cohorte qui précède immédiatement au même âge (observée en 1994). La densité propre à la cohorte la plus jeune (observée en 2003) est perceptiblement moins étalée et les queues de gauche et de droite sont plus minces. En d'autres termes, cette cohorte voit sa littératie s'améliorer à l'extrémité inférieure de la distribution et se détériorer à l'extrémité supérieure. C'est ce dont témoignent un 10^{e} percentile de 223 et un 90^{e} percentile de 363 en 1994 comparativement à des valeurs de 248 et 354 en 2003. Les baisses d'inégalité aux deux extrémités s'annulent presque avec une médiane de 299 en 1994 et de 306 en 2003. On pourrait conclure, par la tendance centrale, que les choses n'ont guère changé et, par les queues de distribution, que les différences sont plus appréciables.

Graphique 2

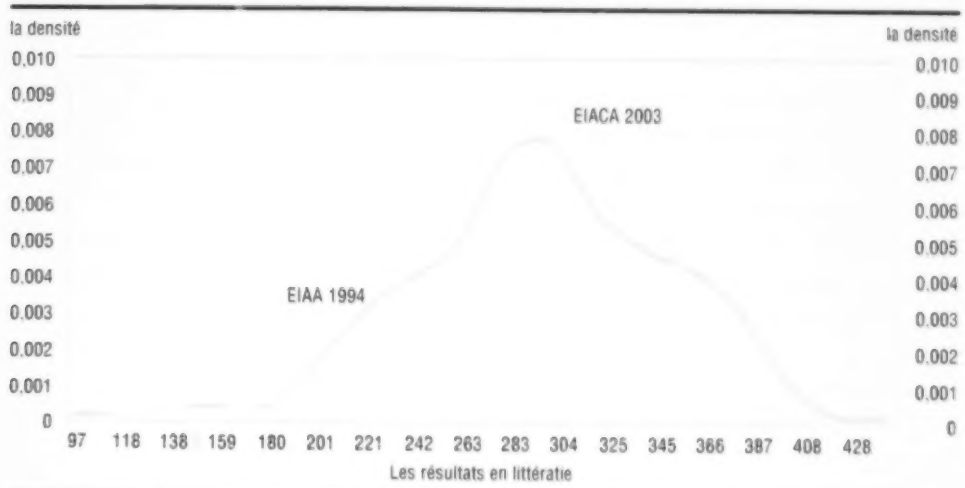
La compréhension de textes schématiques, âgée de 26 à 35 ans



Au graphique 3, nous faisons la même comparaison de valeurs de densité, mais pour les gens âgés de 36 à 45 ans. Le résultat est en gros le même qu'au graphique 2 sauf que l'amélioration relative n'est pas si marquée à l'extrémité inférieure et que la détérioration relative à l'extrémité supérieure l'est décidément plus qu'à cette même graphique. En revanche, le graphique 4 indique que, pour les 46 à 55 ans, les résultats s'améliorent en 2003 dans toute la distribution. Le fait est important, en partie parce que l'observation de variations relatives qui diffèrent entre 1994 et 2003 pour divers groupes donne l'impression de constater quelque chose de bien réel plutôt que de simples différences d'examen entre les deux périodes. Si toutes les estimations de tous les groupes marquent une amélioration à l'extrémité inférieure et une détérioration à l'extrémité supérieure entre 1994 et 2003, l'explication la plus simple en serait que l'examen a changé et que, par conséquent, il donne de meilleurs résultats aux questions faciles qui sont à la base de la plupart des résultats du bas, mais de pires résultats aux questions plus difficiles qui définissent la forme du haut de la distribution. Remarquons enfin que, au graphique 5, il y a amélioration en 2003 pour les 56 à 65 ans, bien que les valeurs de densité se ressemblent de près tout au haut de la distribution entre 1994 et 2003.

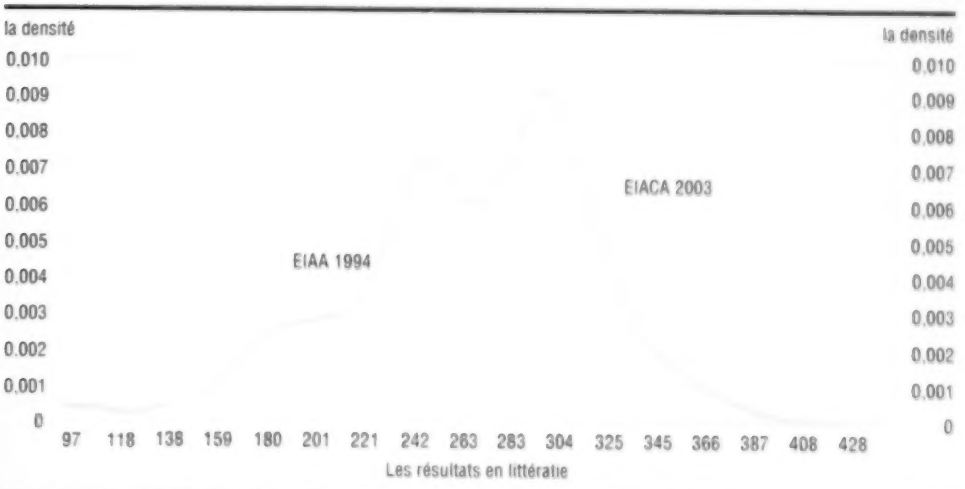
Graphique 3

La compréhension de textes schématiques, âgée de 36 à 45 ans



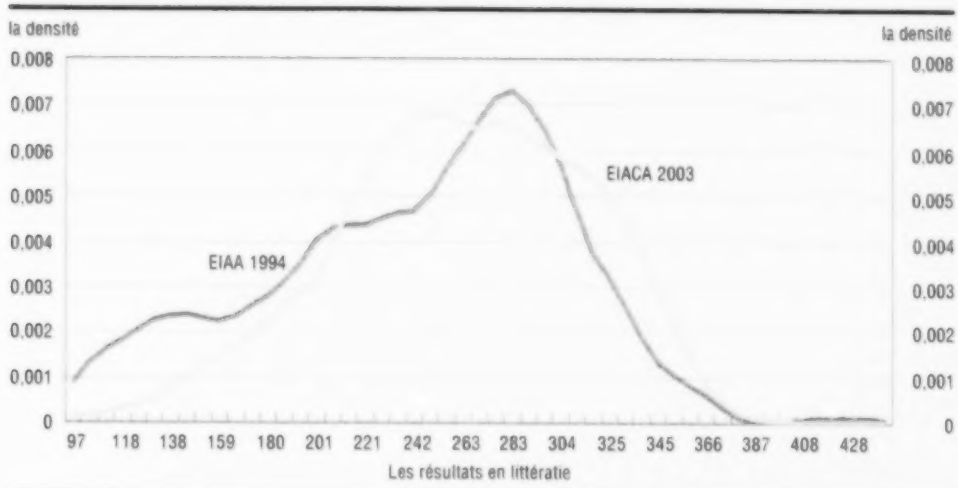
Graphique 4

La compréhension de textes schématiques, âgée de 46 à 55 ans



Graphique 5

La compréhension de textes schématiques, âgée de 56 à 65 ans

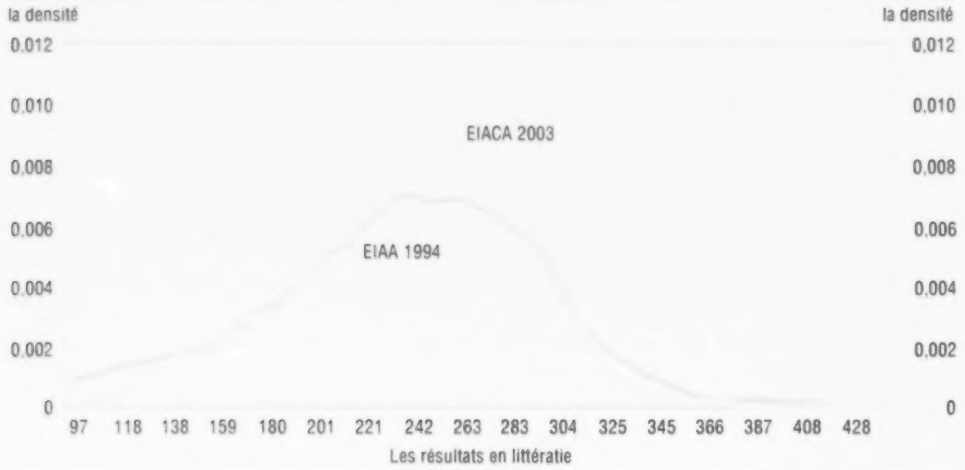


Nous examinons ensuite les variations entre les deux années selon les groupes de scolarisation. Au graphique 6, nous faisons une estimation de densité pour 1994 et 2003 dans le cas de la maîtrise des textes schématiques en visant les gens ayant fait moins que les études secondaires, c'est-à-dire n'ayant pas le diplôme du palier secondaire. L'amélioration à l'extrémité inférieure de la distribution générale se remarque une fois de plus au bas de la distribution de ce groupe et sa valeur est plutôt élevée. Pour ce groupe de décrocheurs, le 10e percentile monte de 145 en 1994 à 175 en 2003. À l'autre extrémité, les différences sont moindres. Si le 95e percentile diminue de 1994 à 2003, le 90e percentile est en réalité le même. Bien sûr, le 90e percentile de la distribution du groupe est relativement bas, correspondant approximativement au 70e percentile de la distribution générale.

Même lorsque, au graphique 7, nous visons les diplômés de l'école secondaire, une détérioration à la queue à droite de la distribution ressort davantage, tout comme une amélioration à la queue de gauche qui le cède nettement à celle du groupe des décrocheurs de l'école secondaire.

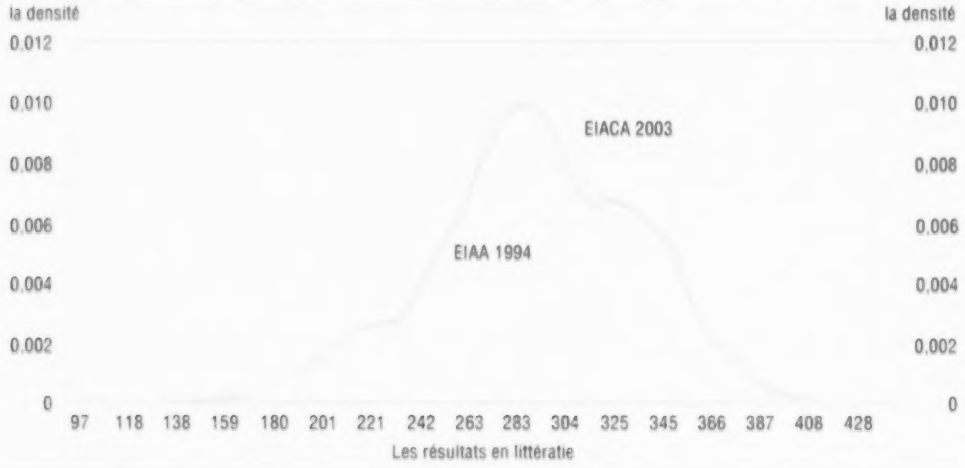
Graphique 6

La compréhension de textes schématiques, moins que des études secondaires



Graphique 7

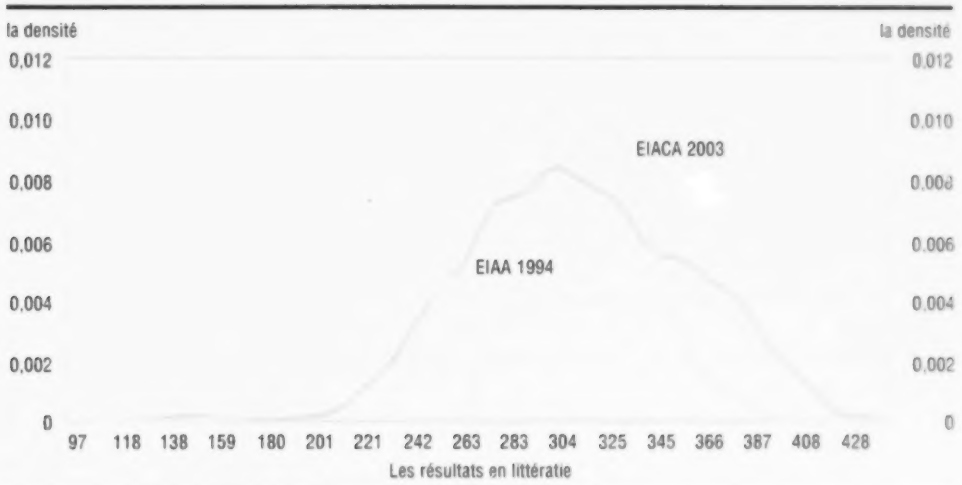
La compréhension de textes schématiques, études secondaires



Aux graphiques 8 et 9 respectivement, on relève pour les groupes « études postsecondaires non universitaires » et « études universitaires » une détérioration à l'extrémité supérieure de la distribution avec, pour le premier de ces groupes, des baisses ailleurs dans la distribution. Pour l'un et l'autre de ces groupes, le 90^{e} percentile diminue environ de 10 % entre les deux années, passant de 373 à 338 pour le groupe « études postsecondaires non universitaires » et de 392 à 359 pour le groupe « études universitaires ». Ainsi, sur le plan des variations intragroupes de cette ventilation selon la scolarisation, l'amélioration constatée au bas de la distribution générale se remarque clairement seulement dans la distribution du groupe des décrocheurs de l'école secondaire et les baisses au haut de la distribution générale s'observent principalement dans les distributions des deux groupes supérieurs.

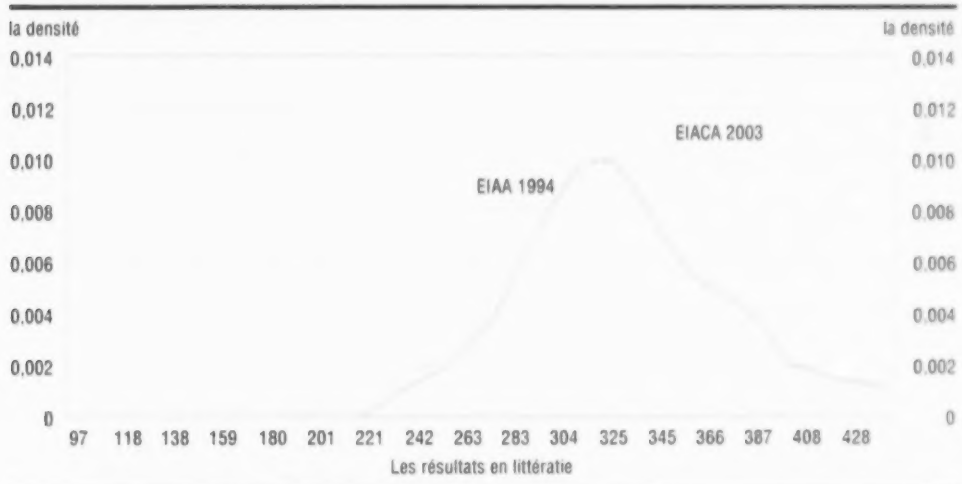
Graphique 8

La compréhension de textes schématiques, études postsecondaires non universitaires



Graphique 9

La compréhension de textes schématiques, université



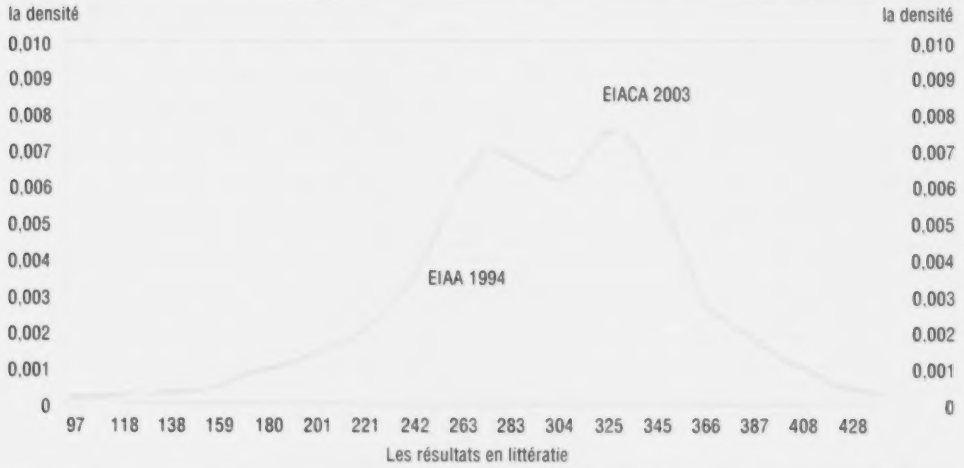
Les différences relevées d'avec la distribution générale pourraient en partie s'expliquer aussi par une évolution de la composition de scolarisation de la population entre 1994 et 2003. Qu'à des cohortes moins scolarisées succèdent des cohortes plus scolarisées a pour effet d'infléchir la distribution selon la scolarisation. Ainsi, la proportion de la population titulaire d'un baccalauréat monte de 17 % en 1994 à 21 % en 2003. La progression a sûrement pour effet d'améliorer la situation à l'extrémité inférieure de la distribution générale. On voit mal cependant comment elle pourrait causer les baisses à l'extrémité supérieure. Si avec la même capacité foncière on est diplômé d'école secondaire dans la cohorte la plus ancienne, mais diplômé d'université dans la cohorte la plus récente, le remplacement de la première par la seconde aurait pour effet une amélioration dans la distribution générale (les études universitaires augmentant la littératie) ou à tout le moins une absence de changement (au cas où cet effet ne

se manifesterait pas). Même le changement au haut de la distribution pour les diplômés d'université n'a probablement rien à voir avec un accroissement du nombre de gens ayant fait des études universitaires. Si ce surcroît de diplômés d'université (qui n'auraient pas fréquenté l'université dans une génération antérieure) sera sans doute d'une capacité inférieure à celle des diplômés d'université dans les deux périodes, c'est l'extrémité inférieure qui serait touchée – et non pas l'extrémité supérieure – dans la distribution du groupe « études universitaires ».

Aux graphiques 10 à 13, nous présentons des estimations d'intensité par année pour des cohortes déterminées de naissances, ainsi que nous les avons définies. À noter le contraste avec les graphiques 2 à 5 où nous examinons les gens d'un même âge dans les deux enquêtes. Les comparaisons des graphiques 2 à 5 sont donc des comparaisons entre cohortes (du même âge) et les comparaisons des graphiques 10 à 13 permettent, elles, de suivre individuellement les cohortes dans le temps.

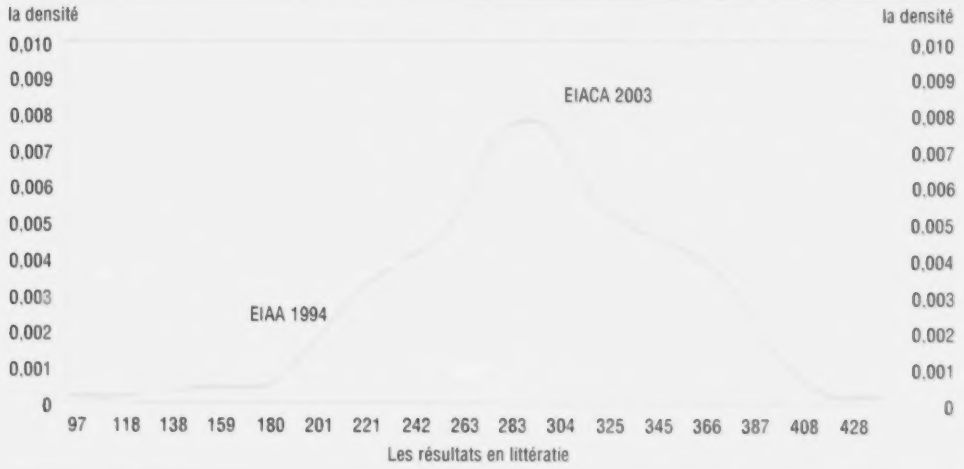
Graphique 10

La compréhension de textes schématiques, âgée de 26 à 35 en 1994



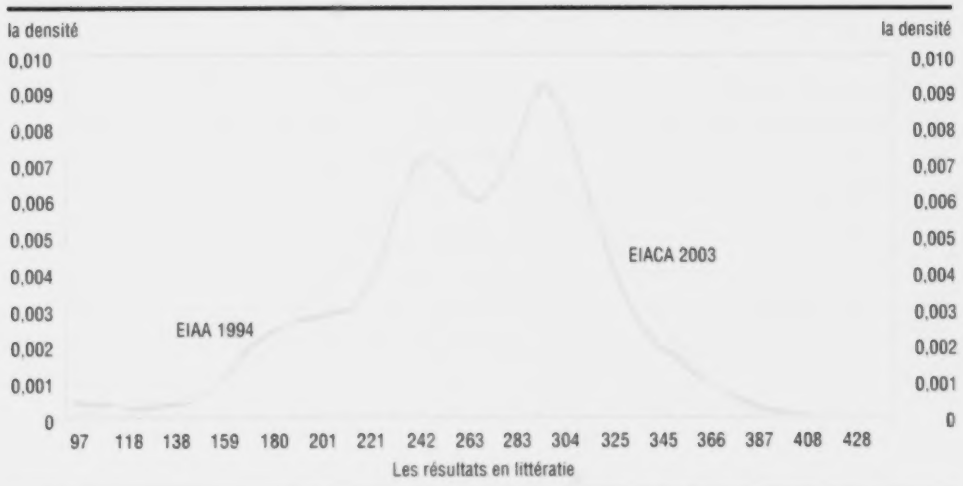
Graphique 11

La compréhension de textes schématiques, âgée de 36 à 45 en 1994



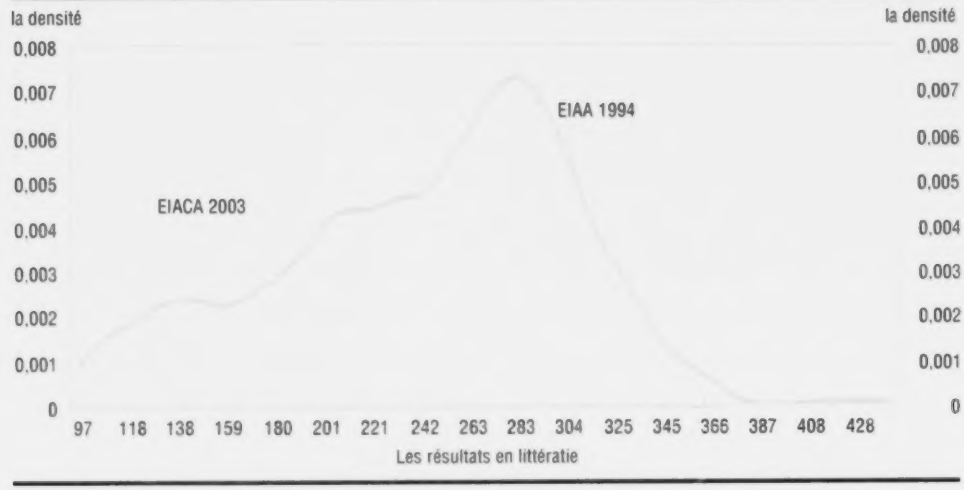
Graphique 12

La compréhension de textes schématiques, âgée de 46 à 55 en 1994



Graphique 13

La compréhension de textes schématiques, âgée de 56 à 65 en 1994

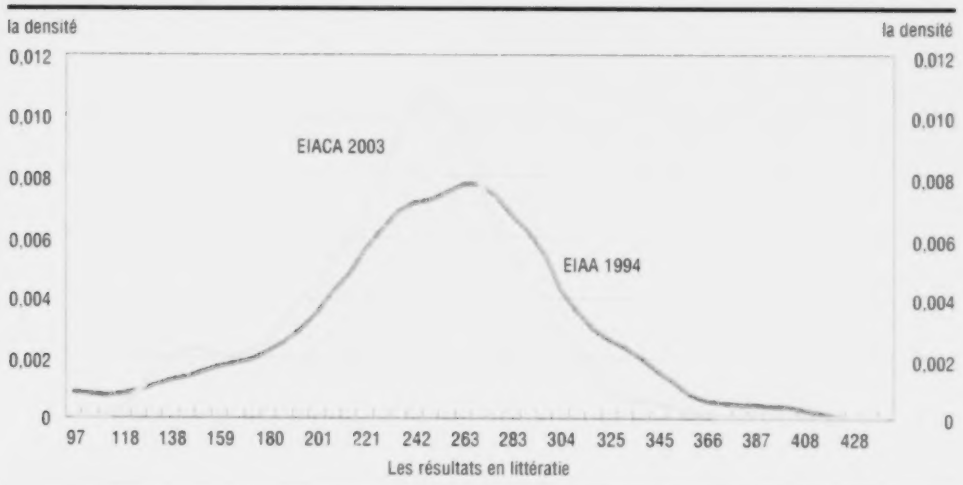


Le graphique 10 livre les estimations relatives à la cohorte âgée de 26 à 35 ans en 1994 et de 35 à 44 ans en 2003. On ne voit guère de différence à l'extrémité inférieure de la distribution pour ces deux années, mais il y a nettement détérioration à l'extrémité supérieure. En fait, le 10e percentile de l'échelle de répartition de la littératie monte de 223 à 232 pour cette cohorte et le 90e percentile tombe de 363 à 344. La tendance est très convergente pour la cohorte âgée de 36 à 45 ans en 1994 au graphique 11; par contre, les valeurs de densité de la cohorte de 46 à 55 ans font voir à tout le moins une amélioration de 1994 à 2003 (graphique 12). Le 10e percentile monte de 187 à 200 dans ce cas et le 90e percentile, de 320 à 328 entre 1994 et 2003. La chose est importante, redisons-le, pour des raisons de comparabilité des données, puisqu'on constate que cette cohorte répond à toutes les questions presque aussi bien (ou peut-être un peu mieux encore) en 2003 qu'en 1994. Les différences de tendances entre cohortes seraient très difficilement justifiables simplement parce qu'on aurait posé, au haut de la distribution, des questions plus difficiles en 2003 qu'en 1994. Une explication plus vraisemblable est que les compétences en littératie se dégradent après les études, mais que le rythme de cette décroissance diminue avec l'âge.

Aux graphiques 14 à 16 enfin, nous présentons les valeurs de densité d'un groupe déterminé cohorte-scolarisation. Là, nous constituons des cohortes plus larges pour disposer d'un nombre suffisant d'observations. Le graphique 14 indique les valeurs de densité du groupe ayant fait moins que les études secondaires. On relève de clairs indices d'une détérioration des compétences en littératie au fil des ans, et ce, à tous les niveaux audessus du 10e percentile environ. La même constatation vaut pour les diplômés de l'école secondaire au graphique 15. Pour les diplômés d'université (voir le graphique 16), les variations sont inégales presque jusqu'à la médiane, mais là encore il y a détérioration nette au-dessus. Il y aurait donc dégradation des compétences supérieures en littératie dans tous les groupes de scolarisation.

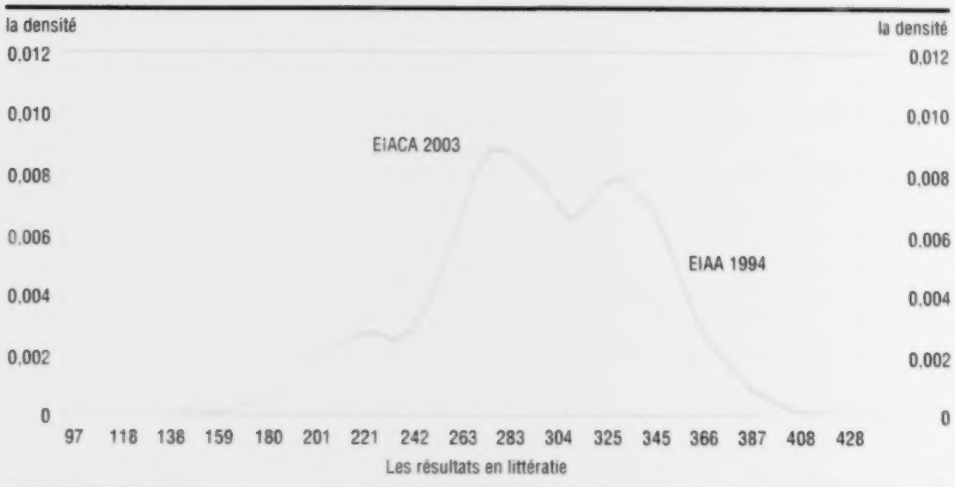
Graphique 14

La compréhension de textes schématiques, moins que des études secondaires, âgée de 26 à 45 en 1994



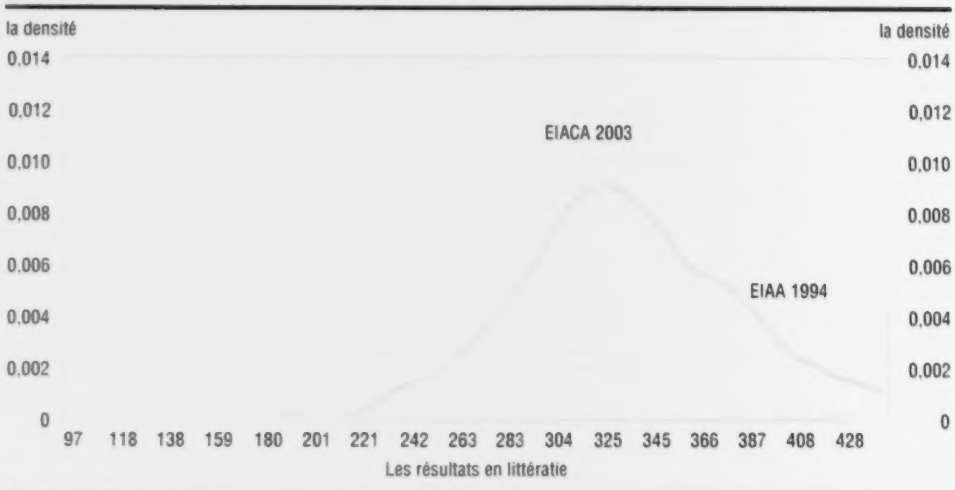
Graphique 15

La compréhension de textes schématiques, études secondaires, âgée de 26 à 45 en 1994



Graphique 16

La compréhension de textes schématiques, université, âgée de 26 à 45 en 1994



Pour étudier encore mieux ces questions d'effets de vieillissement et de cohorte, nous estimons une suite de régressions où nous réunissons les données de l'EIAA 1994 et de l'EIACA 2003 et prévoyons des variables fictives de cohorte dans notre spécification type. La version des données à grande diffusion de l'EIAA 1994 ne comporte pas de variable continue de l'âge, et nous devons rajuster notre spécification type en faisant appel à des variables fictives de l'âge au lieu de variables directes de l'âge et de l'âge au carré. Nous livrons pour cette spécification les résultats d'une estimation simple MCO à la colonne MCO du tableau 4. À noter que, dans toutes les spécifications qui suivent, il y a des variables de l'immigration et de la scolarisation des parents. Dans tous les cas, ces variables se comportent comme dans les estimations antérieures (la faible scolarisation des parents a une incidence négative, mais les autres variables n'ont rien de significatif) et, par souci de concision, nous ne les mentionnons pas. Comme dans la spécification précédente, l'importance des variables de la scolarisation est élevée et significative. L'effet de la scolarisation est plus marqué que dans les tableaux antérieurs. Cela ressort nettement une fois que nous tenons compte des variables fictives de cohorte. Pour l'essentiel, la variable de la scolarisation dans les spécifications antérieures n'appréhendait qu'en partie les effets de cohorte que nous indiquons : les cohortes plus anciennes sont à la fois moins scolarisées et ont un plus haut niveau de littératie, d'où une sous-estimation de l'incidence véritable de la scolarisation.

Tableau 4

Régressions de regroupement avec les effets de cohorte

Variable	MCO	Erreur-type	10e quantile	Erreur-type	Médiane	Erreur-type	90e quantile	Erreur-type
Années de scolarité	0.083***	(0.0067)	0.13**	(0.0052)	0.086***	(0.0031)	0 043***	(0.0028)
Années de scolarité en valeur quadratique	-0.002***	(0.002)	-0.0033***	(0.0002)	-0.0021***	(0.0001)	-0.0009***	(0.0001)
Sexe féminin	-0.0087	(0.0058)	-0.0064	(0.0083)	-0.016	(0.0049)	-0.0062*	(0.0036)
Groupe d'âge								
36 à 45	-0.024*	(0.013)	-0.0074	(0.018)	-0.018	(0.011)	-0.05***	(0.0083)
46 à 55	-0.05**	(0.016)	0.0034	(0.025)	-0.048***	(0.015)	-0.12***	(0.011)
56 à 65	-0.09***	(0.021)	-0.084**	(0.031)	-0.099***	(0.019)	-0.17***	(0.014)
65 et plus	-0.19***	(0.033)	-0.18***	(0.039)	-0.18***	(0.026)	-0.3***	(0.016)
Cohorte 2	0.018*	(0.01)	-0.013	(0.018)	0.018*	(0.0097)	0.055***	(0.0078)
Cohorte 3	0.032**	(0.016)	0.031	(0.025)	0.042**	(0.016)	0.082***	(0.012)
Cohorte 4	0.053**	(0.029)	0.052*	(0.032)	0.055**	(0.023)	0.13***	(0.014)
Cohorte 5	-0.01	(0.033)	-0.053	(0.039)	-0.004	(0.025)	0.11***	(0.015)
Constante	4,97***	(0.054)	4.39***	(0.047)	4.98***	(0.025)	5.48***	(0.021)
Nombre d'observations	14 734	...	14 734	...	14 734	...	14 734	...

... n'ayant pas lieu de figurer

* Différences statistiquement significatives d'avec la valeur nulle aux niveaux de signification de 10 %.

** Différences statistiquement significatives d'avec la valeur nulle aux niveaux de signification de 5 %.

*** Différences statistiquement significatives d'avec la valeur nulle aux niveaux de signification de 1 %.

Nota : Toutes les spécifications comprennent des variables de contrôle pour l'immigration et la scolarisation des parents (données non présentées). La cohorte 1 (catégorie omise) est formée des gens âgés de 26 à 35 ans en 1994; les tranches d'âge correspondantes pour les cohortes 2, 3, 4 et 5 sont 36 à 45, 46 à 55, 56 à 65 et 65 et plus.

Dans les résultats MCO de regroupement pour l'âge et la cohorte, on relève des tendances intéressantes qui correspondent à celles des graphiques. Si nous estimons en particulier notre spécification type à l'aide des données de l'EIACA 2003 et remplaçons les variables de l'âge et de l'âge au carré par les variables fictives des intervalles d'âge, nous dégageons une fois de plus un profil d'âge relativement plat sur la majeure partie de la variation d'âge. Notons en particulier que les deux premières variables fictives (celles des tranches d'âge 36 à 45 et 46 à 55)

ne présentent pas de coefficients élevés ni de différences statistiquement significatives par rapport à la tranche de référence (26 à 35 ans). Pour les deux groupes plus âgés, il y a des effets négatifs et, pour la tranche 65 ans et plus, le coefficient est de -0,16. Cela correspond de près aux tendances estimées à l'aide des variables de l'âge et de l'âge au carré. Avec les variables fictives de cohorte cependant, les effets d'âge présentent des valeurs continues en pente avec une diminution de la littératie moyenne de 2,4 % pour les 36 à 45 ans et de 5 % pour les 46 à 55 ans par rapport au groupe de référence. Par ailleurs, toutes les variables fictives de cohorte sauf la dernière ont des effets positifs significatifs qui s'accroissent avec la cohorte. Ainsi, la cohorte des 56 à 65 ans en 1994 se caractérise par des niveaux de littératie moyenne qui sont supérieurs de 5 % à ceux de la cohorte de référence des 26 à 35 ans en 1994. La cohorte la plus ancienne présente des niveaux qui correspondent en gros à ceux de la cohorte la plus jeune. Dans l'ensemble, l'implication est qu'un profil d'âge de la littératie qui est plat (du moins jusqu'à l'âge de 55 ans) s'explique par une combinaison de perte de littératie avec l'âge et de diminution de la littératie moyenne pour les cohortes plus récentes.

Dans le reste des colonnes du tableau 4, nous étudions ces phénomènes à l'aide de régressions par quantiles sur tout l'échantillon. Nous montrons comment, par exemple, le 10^{e} percentile de l'échelle de répartition de la littératie en fonction des variables incluses évolue au gré des variations marginales des diverses covariables. Ainsi, on peut se reporter au coefficient de la scolarisation pour répondre à la question suivante : comment le 10^{e} percentile de la distribution des résultats de littératie pour les gens ayant 11 ans de scolarité diffère-t-il du percentile correspondant pour les gens ayant 10 ans de scolarité? Plusieurs points ressortent dans ces régressions. Premièrement, l'incidence de la scolarisation diminue selon les quantiles : les années de scolarité ont plus pour effet de modifier les quantiles inférieurs que les quantiles supérieurs. Dans l'examen de ces données, il importe de tenir compte des termes de second ordre, car l'effet de la scolarisation diminue plus rapidement avec la prolongation de la scolarisation dans le quantile le plus bas et prend plus la forme d'une progression linéaire au haut de la distribution. C'est que la scolarisation a bien plus pour effet de changer le bas que le haut de la distribution pour un petit nombre d'années de scolarité, mais qu'elle présente des effets à peu près égaux au haut et au bas pour environ 16 années de scolarité. Deuxièmement, la diminution de la littératie avec l'âge est bien plus prononcée au haut qu'au bas. Comme dans les graphiques, le bas de la distribution est relativement le même pour diverses tranches d'âge, mais pour les groupes plus âgés, le 90^{e} percentile est bien moindre. À regarder le tableau par exemple, on constate que le groupe des 56 à 65 ans présente un 90^{e} percentile qui est inférieur de 17 % à celui du groupe de référence (de 26 à 35 ans) après prise en compte de la scolarisation et de l'instruction des parents. Troisièmement, si les effets de cohorte ressortent dans toute la distribution, ils sont bien plus marqués à l'extrémité supérieure. Ainsi, la cohorte des 56 à 65 ans en 1994 (qui sont nés entre 1929 et 1938) a un 90^{e} percentile supérieur de 13 % à celui du groupe de référence (gens nés entre 1959 et 1968). La question à se poser est de savoir s'il y a eu perte d'efficacité du système scolaire. Que les chiffres fassent voir une certaine amélioration de la littératie parmi les cohortes du bas de la distribution et que les résultats de régression montrent une détérioration, voilà qui est sans doute attribuable au fait que, dans ce dernier cas, on tient compte de la scolarisation et de l'instruction des parents, les cohortes plus jeunes étant plus instruites, tout comme les parents de leurs membres. Collectivement, ces résultats sont peut-être l'indice que l'école réussit moins bien à améliorer la littératie à un niveau déterminé, mais qu'il y a aussi l'avantage bien réel d'une progression de la scolarisation de génération en génération.

Au tableau 5, nous présentons les régressions par quantiles pour les gens ayant le diplôme de l'école secondaire ou moins que les études secondaires et pour les gens ayant fait des études universitaires, le but étant de regarder de plus près les effets liés à la scolarisation. Dans le premier de ces groupes, les effets du vieillissement sont pour ainsi dire inexistants aux quantiles les plus bas, ce qui indique peut-être qu'une littératie très sommaire ne se perd guère avec l'âge.

On est aussi porté à penser que les quantiles les plus bas pour ce groupe de scolarisation se sont en réalité élevés parmi les cohortes. À l'extrémité supérieure de la distribution cependant, les résultats concordent en gros avec les résultats généraux du tableau 4 avec une détérioration marquée avec l'âge et des baisses pour les cohortes récentes. Il paraît donc possible que les variations de l'instruction publique et de la scolarisation secondaire se soient compensées à l'échelle des cohortes avec une amélioration de la littératie au bas de la distribution et une certaine détérioration au haut. Dans le cas des gens ayant fait des études universitaires, la diminution avec l'âge se remarque dans toute la distribution. Il y aurait aussi des baisses parmi les cohortes, mais celles-ci ne sont pas statistiquement bien définies. Il pourrait simplement s'agir d'un problème de taille d'échantillon ou on pourrait penser que les problèmes réels de littératie parmi les générations sont liés aux études postsecondaires non universitaires plutôt qu'aux études universitaires. Il nous faudra pousser l'examen de cette question.

Tableau 5

Régressions de regroupement de quantiles avec les effets de cohorte selon les groupes de scolarisation

	Études secondaires ou moins					
Variable	10e quantile	Erreur-type	Médiane	Erreur-type	90e quantile	Erreur-type
Sexe féminin	0,064***	(0,013)	-0,011	-(0,0049)	-0,0096**	(0,0047)
Groupe d'âge						
36 à 45	0,025	(0,021)	-0,037***	(0,01)	-0,027**	(0,012)
46 à 55	0,019	(0,035)		(0,011)	-0,11***	(0,015)
56 à 65	-0,032	(0,068)	-0,11***	(0,016)	-0,12***	(0,024)
65 et plus	-0,017	(0,12)	-0,28***	(0,024)	-0,24***	(0,027)
Cohorte 2	0,0053	(0,013)	0,038***	(0,01)	0,061***	(0,0094)
Cohorte 3	-0,053	(0,053)	0,015	(0,017)	0,046**	(0,016)
Cohorte 4	-0,23***	(0,068)	0,019	(0,021)	0,044*	(0,025)
Cohorte 5	-0,34***	(0,11)	-0,027	(0,027)	0,042**	(0,024)
Constante	5,45***	(0,027)	5,76***	(0,011)	5,87***	(0,014)
Nombre d'observations	9 209	...	9 209	...	9 209	...
	Études universitaires					
Variable	10e quantile	Erreur-type	Médiane	Erreur-type	90e quantile	Erreur-type
Sexe féminin	-0,021**	(0,01)	-0,0046	(0,0069)	0,006	(0,0084)
Groupe d'âge						
36 à 45	-0,064**	(0,027)	0,0075	(0,012)	-0,09*	(0,05)
46 à 55	-0,074*	(0,042)	-0,021	(0,015)	-0,12***	(0,022)
56 à 65	-0,18***	(0,045)	-0,092***	(0,017)	-0,15***	(0,026)
65 et plus	-0,31***	(0,093)	-0,16***	(0,032)	-0,15***	(0,038)
Cohorte 2	0,012	(0,024)	-0,025	(0,017)	0,0071	(0,009)
Cohorte 3	0,043	(0,032)	0,015	(0,021)	0,018	(0,018)
Cohorte 4	0,085	(0,085)	0,038	(0,039)	0,011	(0,023)
Cohorte 5	0,12	(0,11)	0,0071	(0,041)	-0,12***	(0,035)
Constante	5,73***	(0,032)	5,8***	(0,017)	5,99***	(0,015)
Nombre d'observations	2 163	...	2 163	...	2 163	...

... n'ayant pas lieu de figurer

* Différences statistiquement significatives d'avec la valeur nulle aux niveaux de signification de 10 %.

** Différences statistiquement significatives d'avec la valeur nulle aux niveaux de signification de 5 %.

*** Différences statistiquement significatives d'avec la valeur nulle aux niveaux de signification de 1 %.

Nota : Toutes les spécifications comprennent des variables de contrôle pour l'immigration et la scolarisation des parents (données non présentées). La cohorte 1 (catégorie omise) est formée des gens âgés de 26 à 35 ans en 1994; les tranches d'âge correspondantes pour les cohortes 2, 3, 4 et 5 sont 36 à 45, 46 à 55, 56 à 65 et 65 et plus.

Avant de terminer cet examen des différences d'âge et des questions intergénérationnelles, il importe d'y aller d'une mise en garde. Notre traitement est fondé sur l'idée d'une comparabilité des données d'examen entre l'EIAA 1994 et l'EIACA 2003. Il y a lieu de croire à cette comparabilité (puisque il y a des questions communes et les autres ont été mises en alignement pour que des comparaisons soient possibles, et aussi à cause des différences évoquées de tendances entre les groupes), mais c'est quand même là un éventuel sujet d'inquiétude. Nos résultats seraient renforcés si nous avions accès aux résultats pour la seule proportion de 40 % des questions communes aux deux enquêtes.

4.3 Formation de compétences en littératie dans le milieu de travail après les études

Les résultats présentés à la section précédente indiquent que, au fil des ans, les adultes perdent de leurs compétences en littératie après avoir quitté l'école. Dans la présente section, nous étudions ce que serait un modèle de compétences qui « se perdent à force de ne pas être employées ». Nous regardons en particulier si les gens qui se servent des compétences acquises dans leur emploi maintiennent un plus haut niveau de littératie une fois l'âge pris en compte. Nous réexécutons à cette fin notre spécification de base à l'aide des données de l'EIACA 2003, sans oublier les variables relatives à l'utilisation des compétences au travail et dans la profession exercée. Par les questions sur l'emploi des compétences au travail, on s'enquiert de la fréquence d'accomplissement des tâches de lecture, d'écriture et de calcul. Il y a des questions qui portent respectivement sur cinq tâches de lecture, d'écriture et de calcul. Nous élaborons des variables fictives dont la valeur est l'unité si l'enquêté répond avoir accompli quatre ou cinq des tâches de lecture au moins une fois par semaine. Nous construisons des variables semblables pour les tâches d'écriture et de calcul. Il y a également des variables fictives pour l'accomplissement d'une à trois des tâches visées de lecture, d'écriture et de calcul au moins une fois par semaine. Enfin, nous construisons une variable fictive qui correspond aux gens ayant répondu avoir accompli « rarement » toutes les tâches d'un domaine (celui de la lecture, par exemple). Dans les régressions, la catégorie omise est donc formée des gens qui ont dit avoir accompli certaines des tâches, mais moins d'une fois par semaine.

Nous présentons les résultats de la régression MCO de base avec ces variables de la littératie au travail et une variable de durée d'emploi à la colonne MCO 1 du tableau 6. Pour ne pas encombrer le tableau, nous n'y mentionnons pas les données sur l'immigration et la scolarisation des parents, tout en prenant ces aspects en compte. Les effets ainsi estimés sont fort semblables à ceux qui figurent dans les tableaux antérieurs. L'échantillon comprend tous les gens, qu'ils travaillent ou non. Nous appréhendons les non-travailleurs par la variable fictive de la durée indéterminée d'emploi. L'effet estimé de la durée d'emploi est positif et statistiquement significatif, mais très modeste : si l'occupation d'un emploi se prolonge d'un an, la littératie augmente de 0,06 %. Ainsi, cette dernière ne change pas réellement avec la durée d'emploi. Les variables de littératie au travail indiquent que ceux qui font un usage intensif au travail de leurs compétences en littératie (en déclarant accomplir quatre tâches ou plus dans un groupe au moins une fois par semaine) ont en réalité un meilleur niveau de littératie. L'accomplissement des tâches de lecture au travail est en corrélation étroite avec l'accomplissement des tâches d'écriture, puisque 83 % de ceux qui disent lire rarement précisent écrire aussi peu fréquemment dans leur milieu de travail. Dans l'interprétation des coefficients, il est préférable d'additionner les coefficients de lecture et d'écriture. C'est ainsi que la personne qui accomplit respectivement quatre tâches et plus de lecture et d'écriture par semaine a un meilleur niveau de littératie en moyenne, et ce, dans une proportion approximative de 3,7 %. À l'autre extrême, celui qui écrit ou lit rarement au travail est d'une moindre littératie moyenne dans une proportion d'environ 2 %. Ce sont des effets assez perceptibles pour être dignes de mention, mais ils ne sont pas énormes. S'il y a clairement corrélation entre la littératie d'une

personne et l'utilisation des compétences correspondantes au travail, ceux qui exploitent ces compétences n'ont pas un plus haut niveau de littératie. À noter que nous ignorons en quel sens joue la causalité dans ces résultats. Les gens qui ont un plus haut niveau de littératie auront peut-être plus de chances d'occuper des emplois de haute littératie ou bien de tels emplois pourraient aider les gens à maintenir leurs compétences. Autre possibilité, les deux phénomènes pourraient entrer en jeu l'un et l'autre. L'utilisation des compétences en mathématiques n'est pas autant en corrélation avec les deux autres domaines de compétence au travail que ne le sont ces domaines entre eux. Ainsi, sur le nombre de gens que nous caractérisons comme lisant rarement, 43 % seulement disent aussi utiliser rarement leurs compétences en calcul. Les résultats d'estimation n'indiquent pas une incidence positive sur la maîtrise des textes schématiques d'une utilisation fréquente des compétences en calcul au travail, mais ceux qui ne s'en servent que rarement ont un niveau de littératie moins élevé que le groupe de référence dans une proportion de plus de 3 %.

Tableau 6

Régressions avec les variables de la profession et de l'utilisation au travail des compétences en littératie

Variables	MCO 1	Erreur-type	MCO 2	Erreur-type
Nombre d'années de scolarité	0.047***	(0.0046)	0.0478***	(0.0046)
Nombre d'années de scolarité en valeur quadratique	-0.0009***	(0.0002)	-0.001***	(0.0002)
Sexe féminin	-0.009**	(0.0046)	-0.01**	(0.0049)
Âge	0.0043***	(0.0009)	0.004***	(0.0009)
Âge en valeur quadratique	-0.0075***	(0.001)	-0.0072***	(0.001)
Durée d'emploi	0.0006**	(0.0003)	0.0005**	(0.0003)
Durée d'emploi inconnue	-0.021	(0.0159)	0.025	(0.026)
Utilisation des compétences en littératie au travail				
Lecture : au moins une fois par semaine 4 catégories et plus	0.024**	(0.011)	0.016	(0.011)
Lecture : au moins une fois par semaine 3 catégories et plus	0.019*	(0.011)	0.015	(0.011)
Utilisation rare des compétences en lecture	-0.0045	(0.016)	-0.0047	(0.016)
Écriture : au moins une fois par semaine 4 catégories et plus	0.013	(0.0093)	0.0088	(0.0094)
Écriture : au moins une fois par semaine 3 catégories et plus	0.013	(0.0077)	0.0098	(0.0078)
Utilisation rare des compétences en écriture	-0.023**	(0.01)	-0.021**	(0.01)
Mathématiques : au moins une fois par semaine 4 catégories et plus	-0.01	(0.010)	-0.011	(0.010)
Mathématiques : au moins une fois par semaine 3 catégories et plus	-0.0062	(0.0097)	-0.0064	(0.0098)
Utilisation rare des compétences en mathématiques	-0.035**	(0.015)	-0.033**	(0.015)
Profession				
Membres des Forces armées	...	...	0.044**	(0.016)
Professionnels	...	...	0.006	(0.0077)
Techniciens	...	...	-0.002	(0.0073)
Commis de bureau	...	...	0.006	(0.01)
Travailleurs des services	...	...	0.02**	(0.009)
Travailleurs qualifiés en agriculture	...	...	-0.019	(0.019)
Artisans	...	...	-0.004	(0.0093)
Opérateurs de machines et d'installations	...	...	-0.024**	(0.010)
Simples manœuvres	...	...	-0.052**	(0.014)
Profession inconnue	...	...	-0.067**	(0.027)
Constante	5.16***	(0.04)	5.19***	(0.041)
Nombre d'observations	14 527	...	14 527	...
R^2	0.54	...	0.54	...

... n'ayant pas lieu de figurer

* Différences statistiquement significatives d'avec la valeur nulle aux niveaux de signification de 10 %.

** Différences statistiquement significatives d'avec la valeur nulle aux niveaux de signification de 5 %.

*** Différences statistiquement significatives d'avec la valeur nulle aux niveaux de signification de 1 %.

Nota : Toutes les spécifications comprennent des variables de contrôle pour l'immigration et la scolarisation des parents (données non présentées).

À la colonne MCO 2 du tableau 6, nous ajoutons des variables fictives de la profession. Leur inclusion a une certaine incidence sur les variables de littératie en milieu de travail, mais surtout en accroissant les erreurs-types causées par la collinéarité plutôt que par une nette diminution des effets estimés. Ainsi, le recours fréquent aux compétences en lecture et en écriture au travail implique toujours une augmentation de 2,5 % de la littératie et une utilisation rare, une diminution de 2,1 %. Le groupe professionnel de référence est celui des gestionnaires et, par rapport à lui, les groupes de professionnels, des commis de bureau et des techniciens ne se distinguent pas dans leurs niveaux de littératie. Aspect intéressant, il en va de même des travailleurs agricoles qualifiés, des artisans et des gens de métiers (tous ces résultats sont après prise en compte de la scolarisation). En revanche, les travailleurs des services, les opérateurs de machines et les simples manœuvres ont tous un niveau de littératie moins élevé. Là encore, les effets sont loin d'être considérables. Ainsi, la différence de littératie moyenne entre le gestionnaire et le simple manœuvre est de 5 %. Cette valeur est à comparer à une différence de plus de 30 % entre les gens comptant respectivement 12 et 16 ans de scolarité.

Comme dernière vérification de l'importance de l'utilisation au travail des compétences en littératie, nous avons réexécuté les spécifications regroupées, mais seulement pour les groupes des professionnels, des techniciens et des commis de bureau en posant qu'il s'agit de professions à forte utilisation de compétences. Si l'appartenance à de telles professions permet le maintien des compétences en littératie, nous nous attendrions à ce que le profil d'âge négatif estimé ait une pente moins raide pour ce groupe. Lorsque nous effectuons l'estimation, les coefficients d'âge dégagés sont fort semblables à ceux de l'ensemble des gens à la colonne MCO du tableau 4, et ce peut-être par « glissement » des échantillons (les gens qui occupent les mêmes professions en 1994 ne sont pas nécessairement les mêmes qu'en 2003), mais cet effet est probablement faible au regard de la force probante des résultats. Dans l'ensemble, les indications sont inégales quant à l'incidence de l'utilisation au travail des compétences en littératie. La corrélation est nette entre cette utilisation et les niveaux de littératie moyenne, mais elle demeure modeste par rapport à l'incidence de la scolarisation. Disons enfin que l'appartenance à une profession de forte littératie ne semble rien empêcher de la perte de compétences que subissent les membres de cet échantillon avec l'âge.

5. Incidence de la littératie sur les gains

5.1 Un simple cadre théorique

Nous examinerons maintenant l'importance utilitaire de la littératie sur les gains. Nous avons vu à la section précédente que la littératie est principalement déterminée par la scolarisation. On peut d'emblée penser que la rétribution couramment estimée de la scolarisation correspond en réalité à la rétribution de la littératie. Tant pour représenter cet aspect que pour dresser un cadre d'interprétation des régressions de rémunération qui suivent, nous commençons par présenter un simple modèle hédoniste de la formation des gains. L'examen qui suit est fondé sur Green et Riddell (2003).

Posons que chaque travailleur peut posséder une diversité de compétences, chacune à un degré variable (y compris dans une mesure nulle). Pour simplifier notre exposé, nous pourrions parler de trois compétences. Les gains individuels sont fonction des compétences que possède et exploite l'individu :

$$1)\ E_i = f\ (\ G_i^1,\ G_i^2,\ G_i^3\ ) + \varepsilon_i\ ,$$

où E_i est les gains de l'individu i dans l'année de l'échantillon, G_i^k la quantité de compétence k que i vend sur le marché et ε_i un terme de perturbation indépendant de ces compétences. Ce terme se trouverait à appréhender, pensons-nous, soit des caractéristiques individuelles indépendantes des niveaux de compétence, soit une erreur de mesure des gains. Nous interprétons la fonction f(.) comme une fonction de formation de rémunération qui vient en définitive d'une fonction générale de production décomposable en apports autres que celui des compétences. En caractérisant la fonction f(.), nous pouvons voir l'importance des diverses compétences en question et de leur interaction en production. Pour bien focaliser les idées, nous pouvons considérer que G^1 correspond aux compétences cognitives que mesurent les examens en littératie, G^2 aux autres compétences cognitives (peut-être manuelles) qui, non appréhendées dans de tels examens, peuvent s'acquérir par l'expérience professionnelle et G^3 des caractéristiques non cognitives comme la persévérance qui peut en partie s'acquérir par la scolarisation.

En nous fondant sur 1), nous pouvons élaborer un jeu de fonctions des prix des compétences donné par

$$2)\ r_k = \frac{\partial f}{\partial G^k}\ (G_i^1,\ G_i^2,\ G_i^3),\qquad k = 1,\ldots,3$$

À noter que les prix peuvent varier selon la gamme complète de caractéristiques que vend l'individu. Notre but est de caractériser ce jeu de fonctions des prix des compétences. Cela fait, nous connaîtrons l'importance relative des diverses compétences en production et saurons aussi si elles sont alors en complémentarité ou en substituabilité.

Il serait relativement simple de caractériser 1) ou 2) si nous pouvions observer les compétences G_i^k. Bien sûr, cela est normalement impossible. Ce que nous observons, c'est certains des apports à la formation des compétences. Pour voir comment ces apports s'inscrivent dans notre cadre, considérons un jeu de fonctions de production pour la formation des compétences :

$$3)\ G_i^k = h_k\ (\ yrs_i,\ \exp_i,\ \theta_i)\ .$$

où h désigne la nature des caractéristiques et où yrs correspond aux années de scolarité officielle, exp aux années d'activité sur le marché du travail et θ à un vecteur de capacités innées. À noter que nous distinguons les capacités (qui sont innées) des compétences (qui peuvent s'acquérir et trouvent directement leur utilité en production). Le vecteur des capacités θ peut comprendre des éléments cognitifs ou non. Des capacités non cognitives comme la persévérance pourraient servir à la formation des compétences, cognitives ou non.

Si nous n'observons pas directement les G_i^k, nous pouvons obtenir une équation d'estimation en substituant dans 1) les équations données par 3), ce qui mène à une spécification de forme quasi réduite pour les gains annuels sous la forme

$$4)\ E_i = g(\ yrs_i,\ \exp_i,\ \theta_i) + \varepsilon_i\ .$$

Ainsi, nous considérons un modèle hiérarchisé où des covariables qui entrent normalement dans les régressions salariales sont des apports à la formation des compétences et où ces compétences (avec un terme d'erreur indépendant) déterminent complètement le salaire.

Examinons maintenant les dérivées partielles des gains pour chacun des apports de compétences à la production (scolarisation, expérience professionnelle, éléments du vecteur des capacités, etc.). La dérivée partielle liée à un apport x peut ainsi s'exprimer :

$$5)\ \frac{\partial E}{\partial x} = \frac{\partial f}{\partial G^1} * \frac{\partial h_1}{\partial x} + \frac{\partial f}{\partial G^2} * \frac{\partial h_2}{\partial x} + \frac{\partial f}{\partial G^3} * \frac{\partial h_3}{\partial x}\ ,$$

où nous supprimons le i en exposant pour simplifier. Si x correspond aux années de scolarité yrs par exemple, l'équation 5) nous dit que l'effet observé d'une année supplémentaire de scolarité correspond à l'effet d'une année de scolarité de plus sur la formation de chaque caractéristique en multiplication par le prix que commande cette caractéristique. Il ressort de l'équation 5 que, avec une seule mesure des gains et des apports observables à la formation des compétences, nous ne pouvons rien dire de cette formation ni de la combinaison des compétences en production sauf en indiquant qu'une combinaison critique des dérivées du côté droit de 5) est nulle (et donc $\delta E/\delta x = 0$) ou que certaines ne le sont pas. Si nous disposons d'observations individuelles sur une compétence, disons une compétence G^1, il est cependant possible de dire beaucoup plus.

Si G^1 est observé, la fonction des gains en forme quasi réduite devient :

$$6)\ E_i = g^*(G_i^1,\ yrs_i,\ \exp_i,\ \theta_i) + \varepsilon_i\ .$$

La dérivée de cette fonction par rapport à G^1 correspond à la fonction des prix des compétences r_1, mais nous devons maintenant exprimer le prix visé en fonction de yrs_i, $\exp_i$ et θ_i :

$$7)\ r_i = \chi(G_i^1,\ yrs_i,\ \exp_i,\ a_i)\ .$$

Avec la fonction des prix donnée en 7), il est impossible d'entièrement spécifier les interactions de G^1, G^2 et G^3 en production, mais nous pouvons en apprendre davantage à leur sujet. En particulier, la dérivée de r_1 en fonction de l'apport x à la formation des compétences est égale à ce qui suit (là encore, nous supprimons le i en exposant) :

$$8) \quad \frac{\partial r_1}{\partial x} = \frac{\partial r_1}{\partial G^2} * \frac{\partial h_2}{\partial x} + \frac{\partial r_1}{\partial G^3} * \frac{\partial h_3}{\partial x}.$$

Nous pouvons aussi considérer la dérivée de g^* dans l'équation 6) par rapport à x :

$$9) \quad \frac{\partial g^*}{\partial x} = \frac{\partial f}{\partial G^2} * \frac{\partial h_2}{\partial x} + \frac{\partial f}{\partial G^3} * \frac{\partial h_3}{\partial x}.$$

Avec les valeurs observées des dérivées $\delta r_1/\delta x$ et $\delta g^*/\delta x$, nous pouvons imposer des restrictions à la fonction f. Ainsi, que $\delta h_2/\delta x$ et $\delta h_3/\delta x$ figurent dans toutes ces fonctions donne la possibilité d'imposer des restrictions aux fonctions de production pour les compétences non observées en se fondant sur le signe et la signification des valeurs observées. Avec ces restrictions, nous pouvons peut-être aussi mettre des restrictions aux termes $\delta r_1/\delta G^k$. Dans notre cadre, ces derniers termes reflètent les interactions en production des compétences non observées avec G^1.

Par les différences entre les dérivées en 5) et 9), nous pouvons aussi en apprendre sur la formation de G^1. Les dérivées en question (dérivée des gains, par exemple, par rapport à la scolarisation d'abord sans conditionnement et ensuite en conditionnement par G^1) diffèrent par le terme $\delta f/\delta G^1 * \delta h_1/\delta x$. Ainsi, la différence entre ces dérivées observées indique dans quelle mesure le coefficient pour la scolarisation, par exemple, dans une régression type des gains correspond à l'enchaînement des hausses de la scolarisation, de la rémunération et de la formation de compétences cognitives. Comme nous observons G^1 directement, nous pouvons aller plus loin et mieux voir quelle est la formation de G^1 (selon les termes $\delta h_1/\delta x$) par une estimation directe. C'est essentiellement ce que nous avons fait à la section précédente où notre grande conclusion était que la littératie se forme principalement par la scolarisation et se détériore au fil des ans après les études. Ce dernier effet pourrait être compensé dans une certaine mesure (quoique sans doute d'une manière relativement secondaire) par l'exercice d'un emploi où on exploite ses compétences en littératie.

Il convient de noter que, comme l'exprime l'équation 7), le prix d'une compétence pour l'individu sera fonction du vecteur des capacités θ_i. Pour notre cadre, cela implique que les effets des éléments du vecteur θ_i sur r_1 peuvent s'exprimer comme dans l'équation 8), x étant remplacé par l'élément correspondant de θ_i. En d'autres termes, si nous pouvions observer θ_i, nous en apprendrions davantage sur les interactions des diverses compétences en production. Dans une application purement empirique, le fait que $r_1 = \delta E/\delta G^1$ puisse varier avec des inobservables nous suggère d'employer des régressions par quantiles avec la possibilité d'observer les dérivées des gains par rapport aux variables observables aux différentes valeurs des inobservables qui produisent les divers quantiles conditionnels. De plus, si θ_i était un scalaire au lieu d'un vecteur, que les fonctions h() étaient monotones en θ_i et que la fonction f accroissait en mode monotone les compétences alors les quantiles de la distribution des gains, yrs et exp seraient liés à des valeurs croissantes de θ_i en conditionnement par G^1 et nous pourrions établir le signe de $\delta r_1/\delta\theta$ en nous fondant sur la façon dont $\delta E/\delta G^1$ varie au gré de la progression des quantiles conditionnels de gains.

5.2 Résultats d'estimation

Nous présentons les résultats d'estimation des régressions en moyenne à l'aide du logarithme de la rémunération hebdomadaire au tableau 7. Nous opérons les régressions sur un échantillon restreint qui comprend seulement les gens dont les gains hebdomadaires sont de plus de 50 $ ou de moins de 20 000 $, le but étant d'éliminer quelques valeurs extrêmes. Nous pouvons aussi écarter les travailleurs indépendants, leur rémunération ne pouvant s'adapter à un tel modèle sans qu'il soit aussi tenu compte de la rétribution du capital. La colonne MCO 1 livre les résultats d'une régression type avec une variable fictive du sexe féminin, les années de scolarité, l'expérience et l'expérience au carré comme covariables. Les résultats sont extrêmement normaux dans leur ordre de grandeur et leur signe (voir dans Card (1999) une recension des études très abondantes dans ce domaine). À la colonne MCO 2, nous ajoutons notre variable de la littératie moyenne. C'est la simple moyenne des résultats des quatre domaines de littératie. Si nous agissons ainsi, c'est que notre analyse factorielle des résultats nous le suggère. Elle nous dit en effet que, pour l'essentiel, les résultats sont hautement en collinéarité. Nous avons estimé les spécifications où les quatre grands résultats sont inclus séparément. Dans ces estimations, la maîtrise des textes schématiques est statistiquement significative avec un coefficient de 0,0021; la numératie l'est aussi avec un coefficient de 0,0011. Pour la résolution de problèmes et la maîtrise des textes suivis, les coefficients sont moindres, non statistiquement significatifs et en compensation. À noter que ces effets significatifs séparés s'additionnent essentiellement pour donner le coefficient estimé du tableau 7 pour la littératie moyenne. On peut donc penser que la numératie produit des effets distincts de ceux des trois autres domaines et que ceux-ci sont inférieurs à tout ce qu'appréhendent (principalement) les résultats relatifs aux textes schématiques.

Tableau 7
Régressions des gains

	MCO 1	Erreur-type	MCO 2	Erreur-type	VI 1	Erreur-type	IV 2	Erreur-type
Sexe féminin	-0,41***	(0,024)	-0,4***	(0,024)	-0,4***	(0,024)	-0,4***	(0,024)
Nombre d'années de scolarité	0,087***	(0,0042)	0,069***	(0,0047)	0,07***	(0,013)	0,05***	(0,017)
Expérience	0,067***	(0,0035)	0,067***	(0,0035)	0,068***	(0,0041)	0,068***	(0,004)
Expérience en valeur quadratique	-0,0011***	(0,0001)	-0,0011***	(0,0001)	-0,0012***	(0,0001)	-0,0012***	(0,0001)
Littératie moyenne	...	...	0,0026***	(0,0003)	...	...	0,0032***	(0,0007)
Constante	4,8***	(0,077)	4,28***	(0,10)	5,04***	(0,16)	4,34***	(0,10)
Nombre d'observations	7 768	...	7 768	...	7 768	...	7 768	...
R^2	0,38	...	0,4	...	0,38	...	0,39	...

... n'ayant pas lieu de figurer

*** Différences statistiquement significatives d'avec la valeur nulle aux niveaux de signification de 1 %.

En ajoutant la variable de la littératie moyenne à notre régression type des gains, nous réduisons la dérivée du logarithme des gains par rapport à la scolarisation de 0,087 à 0,069. C'est une baisse d'environ 20 %, ce qui indique à la fois que les compétences en littératie tiennent une grande place dans la rétribution de la scolarisation et que cette dernière a une incidence appréciable sur la rémunération par-delà les effets liés à la formation de ces compétences. L'effet est cependant perceptiblement moindre que ce qui est observé dans l'enquête EIAA 1994 (Green et Riddell, 2003). Il se pourrait que la mesure des gains soit plus lacunaire dans ce premier ensemble de données.

Contrairement au coefficient de la scolarisation, le coefficient des variables de l'expérience professionnelle demeure inchangé lorsque nous introduisons la variable de la littératie. L'explication directe en est que l'acquisition de compétences en littératie n'est liée ni à l'âge ni à l'expérience dans les données transversales. Dans le cadre que nous avons décrit, l'expérience n'entre pas dans la fonction de formation de ces compétences et le premier terme du côté droit de l'équation 5 est nul par conséquent. L'implication est que la dérivée de l'expérience est la même, qu'on fasse intervenir ou non la variable de la littératie. Il reste que nous avons vu que l'âge entre bel et bien – de manière négative – dans la fonction de formation des compétences en littératie, mais que son effet est compensé par des effets de cohorte en décroissance dans la succession des générations récentes. Si l'EIAA 1994 comportait une mesure comparable des gains et que nous pouvions estimer une régression par cohortes regroupées comme nous l'avons fait pour la littératie à la section précédente, nous constaterions que la littératie a sûrement une incidence sur ces coefficients estimés. Disons que la constatation type d'un accroissement des gains avec l'âge pour une cohorte déterminée dans une estimation où la variable de la littératie est absente pourrait, avec la présence de cette variable, comporter une pente encore plus raide[9]. L'implication est que tout accroissement de la rémunération avec l'âge pourrait être dû à la progression encore plus appréciable de la rétribution des compétences hors littératie qui sont acquises par le travail, les compétences en littératie étant, elles, en décroissance.

Signalons enfin que l'incidence directe sur les gains des compétences en littératie est importante. Une majoration de 25 points du résultat moyen (ce qui équivaut à la moitié environ d'un écart-type dans la distribution) est d'une incidence qui équivaut à une année de scolarité de plus.

Comme nous l'avons mentionné dans notre traitement théorique, l'estimation peut subir l'influence des valeurs types de biais de variables omises. Il se peut en particulier que le terme d'erreur de la régression contient divers types de capacités en corrélation avec les variables incluses. Nous supposons normalement que les capacités influent sur les choix de scolarisation et les gains, d'où des biais d'estimation. Dans le modèle que nous spécifions, si nous posons que des capacités cognitives inobservées influent seulement sur la formation des compétences cognitives et que les autres capacités non cognitives n'ont pas une telle incidence, la littératie sera sans corrélation avec le terme d'erreur, et il n'y aura pas de problème d'endogénéité à proprement parler. Il reste à résoudre le problème de l'endogénéité possible de la scolarisation. Nous le faisons en nous reportant à la province de résidence au moment de la scolarisation et en tenant compte de ses éléments d'interaction avec l'âge comme variable instrumentale comme nous l'avons fait dans l'estimation de la littératie. Comme dans ce cas, nous tenons aussi compte de la province actuelle de résidence. En fait, toutes les spécifications comprennent des variables de contrôle de la province de résidence actuelle dont nous ne présentons pas les effets au tableau par souci de concision.

Les résultats avec variable instrumentale pour la scolarisation lorsque la variable de la littératie est exclut figurent à la colonne IV 1 du tableau 7. L'introduction de la variable instrumentale diminue un peu le coefficient de la scolarisation, ce qui ferait voir des problèmes d'endogénéité avec l'estimation simple MCO. Nous reprenons l'exercice à la colonne IV 2, mais en incluant aussi l'effet de la littératie moyenne. Nous obtenons un coefficient quelque peu moindre pour la scolarisation que dans l'estimation simple MCO à la colonne MCO 2. Comme dans cette estimation, l'introduction de la variable de la littératie moyenne réduit le coefficient de la scolarisation d'environ 20 %. Le coefficient même de la variable de la littératie s'élève quelque peu, mais il demeure en gros du même ordre de grandeur qu'à la colonne 2. Ainsi, nos principales conclusions ne sont pas du tout touchées par le recours à une variable instrumentale.

Comme nous l'avons indiqué, notre cadre théorique fait voir les avantages de l'application d'un cadre de régression par quantiles. Nous présentons les résultats d'une telle estimation de régression pour les 10e, 25e, 50e, 75e et 90e quantiles au tableau 8. Voici les grandes implications de cette estimation. Disons d'abord que la rétribution tant de la scolarisation que de l'expérience professionnelle est en décroissance selon les quantiles. Le constat d'hétérogénéité de la rétribution de la scolarisation à l'échelle des gains a déjà été dressé par d'autres auteurs. Buchinsky (1997) constate que la rétribution de la scolarisation s'élève le long des quantiles pour tous les groupes d'expérience professionnelle. Arias et coll. (2001) estiment des régressions semblables par quantiles à l'aide de données de description de jumeaux aux États-Unis et de méthodes permettant de résoudre les problèmes d'endogénéité. Dans une estimation sans variable instrumentale, ils constatent que le coefficient de la scolarisation monte du 10e au 50e percentile, mais ne varie pas dans la partie supérieure de la distribution. Lorsqu'ils emploient des variables instrumentales pour l'erreur de mesure et l'« état de jumeau » en correction du biais des capacités, les coefficients qu'ils estiment pour la scolarisation paraissent relativement homogènes dans toute la distribution, mais ne font pas l'objet d'une estimation précise aux extrémités.

Tableau 8

Régressions des gains par quantiles

	10e quantile	Erreur-type	25e quantile	Erreur-type	Médiane	Erreur-type	75e quantile	Erreur-type	90e quantile	Erreur-type
Sexe féminin	-0.48***	(0.048)	-0.46***	(0.032)	-0.36***	(0.020)	-0.36***	(0.021)	-0.36***	(0.032)
Nombre d'années de scolarité	0.079***	(0.011)	0.065***	(0.0065)	0.069***	(0.0041)	0.067***	(0.0042)	0.057***	(0.0069)
Expérience	0.088***	(0.0063)	0.084***	(0.0041)	0.059***	(0.0025)	0.05***	(0.0024)	0.042***	(0.0036)
Expérience en valeur quadratique	-0.0015***	(0.0001)	-0.0015***	(0.0001)	-0.001***	(0.0001)	-0.0008***	(0.0001)	-0.0006***	(0.0001)
Alphabétisation moyenne	0.0022***	(0.0007)	0.0032***	(0.0004)	0.0029***	(0.0003)	0.003***	(0.0003)	0.0028***	(0.0005)
Constante	3.33***	(0.19)	3.67***	(0.12)	4.26***	(0.083)	4.62***	(0.079)	5.12***	(0.13)
Nombre d'observations	7.768	...	7.768	...	7.768	...	7.768	...	7.768	...

... n'ayant pas lieu de figurer

*** Différences statistiquement significatives d'avec la valeur nulle aux niveaux de signification de 1 %.

Peut-être le résultat le plus intéressant au tableau 8 est-il l'absence de variation du coefficient de la variable de la littératie le long des quantiles. Si le coefficient du 10e quantile semble bien moindre que ceux des autres, il n'a rien de statistiquement significatif. Toutefois, avec des régressions aux quantiles immédiatement voisins du 10e (5e et 15e quantiles), nous obtenons des effets estimés de la littératie qui correspondent presque parfaitement à ceux des quantiles supérieurs. Le 10e ressemble un peu à une valeur statistique aberrante. Dans le cadre de notre modèle théorique, cela implique que les compétences cognitives n'entrent pas en interaction avec les autres caractéristiques dans la formation de la rémunération. Ainsi, d'autres caractéristiques comme la beauté (Hamermesh et Biddle, 1994) et les qualités de leadership (Kuhn et Weinberger, 2005) peuvent contribuer à la rémunération individuelle, mais le rapport ne s'améliore pas par l'augmentation des compétences en littératie, lesquelles n'ont rien d'une panacée tant pour les gains que pour la rétribution des autres caractéristiques. En d'autres termes, dans la mesure où les qualités de leadership concourent à la rémunération et peut-être à la productivité, ces résultats ne suggèrent pas aux entreprises de cibler la formation en leadership selon les niveaux de compétences cognitives, puisque des travailleurs aux compétences cognitives différentes recevront les mêmes avantages de cette formation.

Il faut enfin dire que les estimations que nous présentons des effets de la littératie sont identiques à nos constatations avec les données de l'EIAA 1994 (Green et Riddell, 2003). Dans les deux cas et avec les spécifications types, le coefficient est de 0,0029 pour la littératie moyenne dans une régression des gains en médiane. Nous y voyons un effet du hasard dans une certaine mesure, puisque les mesures des gains ne sont pas les mêmes dans les deux études et que, pour d'autres variables, on y estime des effets différents. Les grandes tendances estimées avec les données de l'EIAA 1994 subsistent néanmoins, dont la taille de l'effet de la littératie et son absence de variabilité selon les quantiles. Comme dans cette étude, nous concluons que la littératie est un grand facteur déterminant de la rémunération, mais qu'une grande partie de la variation des gains est à mettre au compte d'autres facteurs.

6. Conclusion

Dans cette étude, nous avons considéré la littératie autant dans son importance foncière que dans son incidence sur la rémunération. Nous constatons qu'elle est d'une répartition bien plus égale que le revenu et que le degré d'inégalité dans la distribution de la littératie a diminué entre les années d'enquête 1994 et 2003. Un examen plus attentif révèle que la littératie s'est améliorée à l'extrémité inférieure de la distribution, mais qu'elle s'est détériorée à l'extrémité supérieure. Les décrocheurs de l'école secondaire auraient un meilleur niveau de littératie, mais ce résultat serait surtout attribuable à la progression de la scolarisation de la population. Les baisses au haut de la distribution se remarquent pour les diplômés des écoles secondaires, des universités et des établissements postsecondaires non universitaires. Nous constatons aussi que, pour la succession des cohortes de naissances, la littératie se dégrade au haut de la distribution et que le système éducatif donne de meilleurs résultats en littératie au bas de la distribution et des résultats plus médiocres au haut. On peut aussi voir que la littératie tend à décroître avec l'âge après les études et que la scolarisation est le grand moteur de la littératie. La scolarisation des parents a étroitement à voir avec la littératie, mais d'une manière intéressante : les parents qui ont décroché de l'école secondaire ont une forte influence négative, mais au-dessus du groupe des diplômés du palier secondaire, la scolarisation parentale n'ajoute rien comme effet. On est porté à croire à une certaine corrélation entre la littératie et l'utilisation des compétences correspondantes au travail, d'où l'impression (avec le phénomène de la décroissance avec l'âge) que le modèle postscolaire en est un de « compétences qui se dégradent à force de ne pas être employées ». Il reste que les effets en question ne sont pas importants.

Dans la formation de la rémunération, la littératie tient une grande place. Selon nos estimations par les moindres carrés ordinaires, une augmentation de 25 points de la littératie moyenne (ce qui équivaut à environ la moitié d'un écart-type dans la distribution des résultats de littératie) est d'une incidence qui équivaut à une année supplémentaire de scolarité. Lorsque nous introduisons la variable de la littératie dans une régression type des gains, le coefficient de la variable de la scolarisation diminue d'environ 20 %, ce qui implique que le cinquième environ de l'incidence ordinairement mesurée de la scolarisation sur la rémunération se présente uniquement parce que la scolarisation améliore la littératie. Nos résultats pour les variables instrumentales paraissent conférer un rôle plus important à la littératie au moment de mesurer la rétribution de la scolarisation, mais il faudra pousser l'investigation pour le savoir. Que le coefficient de la scolarisation demeure fort et positif indique que celle-ci influe aussi sur les gains par son incidence sur des caractéristiques non cognitives comme les qualités de leadership et l'assiduité au travail.

Référence

Aris, O., K.F. Hallock et W. Sosa-Escudero (2001). "Individual Heterogeneity in the Returns to Schooling: Instrumental Variables Quantile Regression Using Twins Data," Empirical Economics (26), pp. 7-40.

Buchinsky, M. (1997). "Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research," The Journal of Human Resources, 33(1), pp. 88-126.

Card, D. (1999). "The Causal Effect of Education on Earnings," dans O. Ashenfelter et D. Card (eds.) Handbook of Labor Economics, Vol. IIIa. Amsterdam: Elsevier.

Crompton, S (1996). "The Marginally Literate Workforce," Perspectives on Labour and Income, Statistique Canada, été.

Frenette, M., D.A. Green et K. Milligan (2006). "Nouvelle analyse des tendances récentes de l'inégalité du revenue après impôt au Canada au moyen des données de recensement." University of British Columbia et Statistique Canada.

Green, D.A. et Riddell, W.C. "Literacy and Earnings: An Investigation of the Interaction of

Cognitive and Unobserved Skills in Earnings Generation," Labour Economics 10 (Avril 2003) 165-84.

Hamermesh, D.S. et J.E. Biddle (1994). "Beauty and the Labor Market," American Economic Review, 84 (No. 5) pp. 1174-94.

Kuhn, P. et C. Weinberger (2005). "Leadership Skills and Labor Market Outcomes," Journal of Labor Economics, 23, Juillet, 395-436.

Sen, A. (1999). Development as Freedom. New York: Anchor Books.

Annexe A

Définitions de la scolarisation dans l'EIAA 1994, l'EIACA 2003 et l'EPA

Dans cette annexe, nous examinons les différentes questions sur la scolarisation dans l'EIAA 1994 et l'EIACA 2003 et les comparons aux réponses correspondantes dans les reprises de 1994 et 2003 de l'Enquête sur la population active (EPA).

Les questions posées sur la scolarisation dans l'EIAA 1994 et l'EIACA 2003 se ressemblent sans être identiques. Dans l'une et l'autre de ces enquêtes, on s'enquiert du nombre d'années de fréquentation scolaire et une valeur nulle est possible. Dans l'EIACA 2003, on demande ensuite à l'enquêté s'il a terminé ses études secondaires soit directement soit par réussite à un examen d'équivalence. Dans l'EIAA 1994, cette dernière question n'est pas posée. Les deux enquêtes comprennent la question suivante : « Quel est le niveau de scolarité le plus élevé que vous avez complété? » Les éléments possibles de réponse varient cependant selon ces enquêtes. Dans l'EIAA 1994, il y a plus de catégories possibles en deçà du diplôme d'études secondaires, mais dans l'EIACA 2003, il y en a plus entre le palier secondaire et le palier universitaire (5 comparativement à 1 dans l'EIAA 1994). Cette dernière divergence rend possibles des différences de traitement des gens qui obtiennent un certificat de métiers ou d'apprentissage. Ces catégories sont directement énumérées dans la question de l'EIACA 2003. Même les décrocheurs de l'école secondaire qui ont obtenu d'autres certificats peuvent déclarer des résultats supérieurs de scolarisation. Avec la liste plus courte de l'EIAA 1994 où la catégorie correspondante est la très générale rubrique des études postsecondaires non universitaires, on pourrait imaginer que les répondants négligent de mentionner les programmes d'apprentissage, par exemple.

Nous avons expérimenté plusieurs définitions des catégories de scolarisation à l'aide des questions de l'EIACA 2003. Nous nous employons à apparier l'EIAA 1994 avec l'EIACA 2003 plutôt que l'inverse, car les questions et les catégories de scolarisation sont plus fines dans la seconde de ces enquêtes. Notre choix de catégories n'influe guère sur les résultats que nous présentons dans le corps de notre étude. Nous avons opté pour une classification où nous rangeons dans la catégorie « moins que les études secondaires », les gens qui ont répondu ne pas avoir terminé l'école secondaire (à la question directe à ce sujet) et qui indiquent que leur plus haut niveau de scolarité est « moins que les études secondaires », le diplôme d'études secondaires ou des études postsecondaires incomplètes qui n'ont mené ni à un certificat ni à un grade. Nous supposons que ces non-diplômés de l'école secondaire n'auraient pas répondu que leur plus haut niveau de scolarité était les « études secondaires non universitaires » au questionnaire de 1994. Nous caractérisons les gens qui offrent les mêmes réponses à la question sur le plus haut niveau de scolarité, mais qui indiquent aussi être diplômés de l'école secondaire en réponse à la question précédemment posée. La catégorie « études collégiales » comprend toutes les catégories qui, à la question sur le plus haut niveau de scolarité, dépassent les études

postsecondaires incomplètes, mais sans atteindre le baccalauréat. Nous réunissons dans la catégorie « études universitaires » tous les grades d'université, c'est-à-dire le baccalauréat et les grades supérieurs.

Cette caractérisation permet entre les deux ensembles de données une comparaison où chaque entrée aux colonnes EIAA et EIACA indique la proportion de la catégorie correspondante dans l'année.

Tableau A.1
Comparaisons de catégorie d'éducation : EIAA et EIACA

	EIAA 1994	EIACA 2003	Différence
		pourcentage	
Moins que les études secondaires	36.6	25	-11.6
Études secondaires	30.1	28	-2.1
Études postsecondaires non-universitaires	16.8	26.2	9.4
Études universitaire	16.5	20.9	4.4

Comme point de comparaison, nous avons établi les proportions des catégories correspondantes pour 1994 et 2003 dans l'Enquête sur la population active :

Tableau A.2
Comparaisons de catégorie d'éducation : EPA

	EPA 1994	EPA 2003	Différence
		pourcentage	
Moins que les études secondaires	31.8	22.2	-9.6
Études secondaires	26.2	26.4	0.2
Études postsecondaires non-universitaires	26.9	31.3	4.4
Études universitaire	15	20.2	5.2

Les tendances générales sont semblables avec une décroissance en proportion dans la catégorie des études secondaires et une croissance dans les deux catégories postsecondaires. Les proportions de la catégorie universitaire sont aussi fort semblables chaque année entre les deux ensembles de données, mais l'EPA majore la proportion dans la catégorie des études postsecondaires incomplètes, plus particulièrement dans la première de ces périodes, et les diminue en conséquence dans les deux catégories inférieures. Nous en tirons deux conclusions. D'abord, on ne voit pas clairement quel ensemble de données (celui des enquêtes sur la littératie ou de l'EPA) devrait être considéré comme plus sûr. Il convient de noter en particulier que l'EPA comprend les réponses des autres membres du ménage sur la scolarisation de la personne visée s'il est impossible de prendre contact avec cette personne. Nous ignorons combien d'observations sont touchées par cette réponse par procuration, tout comme les éventuelles erreurs de mesure qui en découlent. Dans les enquêtes sur la littératie par ailleurs, la réponse par procuration ne pose aucun problème, puisque la personne visée doit être directement jointe pour subir l'examen. Il peut par contre y avoir des problèmes de représentativité par rapport à l'EPA, puisque cette enquête renseigne au moins quelque peu sur les gens difficiles à joindre directement. En second lieu, les tendances générales se ressemblent entre les deux ensembles de données et les données de l'EIAA paraissent devoir dégager un tableau fidèle de l'incidence de la variation temporelle de la composition de la population dans sa scolarisation.

Annexe B

Résultats du premier degré pour les années de scolarité

Tableau B.1
Régressions du logarithme de la littératie

Variable	MCO	Erreur-type
Âge	0,26***	(0,015)
Âge en valeur quadratique	-0,27***	(0,014)
Scolarisation de la mère		
Moins que les études secondaires	-1,01***	(0,13)
Études postsecondaires incomplètes	0,54***	(0,14)
Baccalauréat ou plus	0,47*	(0,26)
Non-déclaration	-1,56***	(0,20)
Scolarisation du père		
Moins que les études secondaires	-0,77***	(0,13)
Études postsecondaires incomplètes	0,5***	(0,18)
Baccalauréat ou plus	1,31***	(0,19)
Non-déclaration	-1,59***	(0,19)
Mère immigrante	0,44***	(0,17)
Père immigrant	0,49***	(0,16)
Constante	8,68***	(0,40)
Nombre d'observations	13 868	...
R^2	0,23	...

... n'ayant pas lieu de figurer

* Différences statistiquement significatives aux niveaux de 10 %.

*** Différences statistiquement significatives aux niveaux de 1 %.

Nota : La spécification comporte un jeu complet de variables fictives de la province actuelle de résidence et de la province de résidence à l'époque de la fréquentation de l'école secondaire avec tous les éléments d'interaction de ces derniers éléments avec l'âge.

Notes en fin de texte

1. Les autres pays qui ont participé à ce premier cycle sont les Bermudes, l'Italie, le Mexique (État de Nuevo Leon), la Norvège, la Suisse et les États-Unis. Quant à l'EIAA, elle a eu lieu dans plus d'une vingtaine de pays pendant la période 1994 à 1998.
2. Nous avons tronqué l'estimation de la densité à 96 pour conserver la partie principale. Nous éliminons seulement cinq observations, mais à cause du lissage propre aux estimations de densité noyau, la distribution aurait été visuellement dominée par une longue queue mince à gauche. Dans cette estimation, nous employons les valeurs implicites de Stata 7 pour les paramètres de représentation noyau et de lissage.
3. Nous parvenons à cette valeur en prenant la première dérivée du profil d'âge et en évaluant à l'âge de 30 ans. Dans ce profil quadratique, la première dérivée correspond au coefficient des variables linéaires de l'âge, plus deux fois le coefficient de la variable de l'âge au carré en multiplication par 30.
4. Le calcul se fait comme nous le décrivons à la note 3.
5. À noter que le nombre d'observations est moindre maintenant, car nous avons retranché les observations pour lesquelles la province de fréquentation de l'école secondaire est indéterminée.
6. La variable de test correspondant à l'hypothèse commune selon laquelle les variables fictives de la province de résidence et leurs éléments d'interaction avec l'âge sont significativement différents de zéro présente une distribution F (22 13821) et une valeur p de 0,009.
7. Cette conclusion ne se vérifierait pas si la province de résidence à l'époque de la fréquentation de l'école secondaire influait directement sur les résultats de littératie. Tel serait peut-être le cas si les gens qui résident dans des provinces riches en ressources à l'époque de cette fréquentation sous-investissaient à dessein dans leur littératie, sachant d'avance qu'ils vont décrocher et prendre un emploi qui n'exige pas qu'ils aient des compétences en littératie lorsqu'ils atteindront l'âge de l'emploi. Nous devons supposer pour l'essentiel qu'une telle relation n'existe pas, du moins après prise en compte de la province actuelle de résidence.
8. Nous définissons l'expérience par l'expression type de Mincer (expérience = âge – nombre d'années de scolarité - 6). Le coefficient de la variable de l'expérience dans notre régression MCO la plus simple (avec seulement la variable fictive du sexe féminin, la scolarisation, la scolarisation en valeur quadratique, l'expérience et l'expérience en valeur quadratique) est de 0.00051 avec une erreur-type de 0.00045. Le coefficient de cette même variable au carré est de -0.000073 avec une erreur-type de 0.0000083.
9. Nous ne pouvons démontrer ce résultat, mais il s'impose lorsque nous regardons la relation entre l'alphabétisation et l'âge en combinaison avec les valeurs types de biais de variables omises.